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 multi-primer high throughput sequencing (HTS) metabarcoding method and associated analytical workflow for a rapid, intensive, high-volume survey of pest species compositions. We demonstrate our method using the taxonomically challenging *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. Our multi-primer metabarcoding HTS amplicon approach simultaneously process high volumes of whitefly individuals, with efficiency to detect rare (i.e., 1%) test-species, while our improved whitefly primers for metabarcoding also detected beneficial hymenopteran parasitoid species from whitefly nymphs. Field-testing our redesigned *Bemisia* metabarcoding primer sets 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 host crops and sampling techniques that target either nymphs or adults. Our multi-primer HTS metabarcoding method incorporated two over-lapping amplicons of 472bp and 518bp that spanned the entire 657bp 3'

barcoding region for *Bemisia,* and is particularly suitable to molecular diagnostic surveys of this highly cryptic insect pest species complex that also typically exhibited high population densities in heavy crop infestation episodes. 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; multi-primer molecular diagnostics; Aphelinidae; parasitoids; metabarcoding

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 c oxidase subunit I (mtCOI) partial gene 5' (N) terminal region (i.e., typically one of 648 bp, 650 bp, or 658 bp) in 'DNA barcoding' via the Sanger sequencing method has contributed to molecular diagnostics of species (e.g., Hebert, Cywinska, Ball, & DeWaard, 2003; Ward, Zemlak, Innes, Last, & Hebert, 2005; Berry et al., 2004; Boykin et al., 2007; Dinsdale, Cook, Riginos, Buckley, & De Barro, 2010), and provided insights into the potential species diversity where cryptic species may co-exist (e.g., Hebert et al., 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 34 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, Shirima, 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 answer questions around species diversity and evolutionary interactions can be challenging.

The HTS platform (e.g., Illumina's iSeq100, MiniSeq, MiSeq, NextSeq 550, NextSeq 1000 & 2000, NovaSeq 6000; Oxford Nanopore Technologies (ONT) range i.e., Flongle, MinION, GridION, and PromethION; PACBIO's Single Molecule, Real-Time (SMRT) Sequencing and Sequel System/Sequel II System/Sequel IIe System) offers a range of scalable solutions (i.e., to best suit the research institute's and target country's infrastructure) for surveillance of insect species (e.g., see Hebert et al. 2018; Srivanthsan et al. 2021), including cryptic species that reach high population densities, provided that methods are sufficiently flexible to

93 accommodate study designs tailored to address targeted research questions. For example, understanding
94 host-parasitoid relationships, cryptic species diversity, host-plant relationships, and identifying keystone
95 species are fundamental ecological questions common to diverse ecosystems. Addressing these ecological and
96 evolutionary questions can be challenging and requires scaling up of the data collection processes. HTS
97 platforms represent a solution when combined with appropriate species diversity databases, flexible
98 bioinformatics and associated analytical pipelines, and optimised molecular methods.

99 The metabarcoding approach that incorporates the power of HTS platform and the utility of the DNA
100 barcoding gene region aims to provide species identification at both population and community levels (i.e., not
101 aimed at characterising all known mtCOI haplotypes at the individual level). In the cryptic *Bemisia tabaci*
102 species complex however, molecular diagnostics is particularly susceptible to sub-optimal primer efficacies
103 (Elfekih, Tay, et al., 2018; Mugerwa et al., 2018; Shatters, Powell, Boykin, Liansheng, & McKenzie, 2009; Tay,
104 Elfekih, Court, et al., 2017). Partial or inconsistent Polymerase Chain Reaction (PCR) amplification, especially
105 when targeting co-extracted DNA of individual cryptic species that may be found together, can lead to
106 inaccurate conclusion of species composition within the surveyed agroecosystems. For example, a minor
107 species may become over-represented within a population if co-amplified within a pool of major species but
108 enjoys superior primer efficacies compared with the major species. Also, given the time, effort and cost
109 necessarily associated with the preparation of HTS DNA libraries, especially where accurate input genomic
110 DNA concentration is needed, metabarcoding should incorporate a 'back-up' step. Such a step may involve a
111 second set of primers, ensuring the amplification of the same population, should the first set of primers fail
112 and thereby avoiding a total PCR failure. Such a back-up design also enables capturing of species that are
113 missed by the first primer pairs, and vice versa, or where longer DNA region is needed to improve molecular
114 diagnostics of cryptic species complex. Provided that the primer pairs developed were sufficiently robust,
115 within a pest insect species complex such as in the whitefly cryptic species, beneficial insects such as parasitoid
116 species can also be investigated with appropriate sampling of pest-insect life stages (e.g., whitefly nymphs).

117 Here, we describe the HTS metabarcoding method and primers developed for surveying the *Bemisia*
118 cassava whitefly, related *B. tabaci* cryptic species and associated parasitoid species complex, and the workflow
119 for processing the large volume of HTS amplicon sequence reads generated. Our goal was to develop a multi-
120 primer metabarcoding technique that enabled us to determine the dominant *Bemisia* species in any cassava

121 field quantitatively, using a robust sample size in terms of numbers of individuals. We provide a verification of
122 primer efficacies and both successful and failed species delimitation rates of the multi-primer method, and
123 confirmed the method and primer efficacies via Sanger sequencing of individual field-sampled novel whiteflies
124 species. Our samples were collected from geographically diverse cassava cropping landscapes in Uganda,
125 Malawi and Tanzania, so test the limitations and usefulness of this approach especially for countries with
126 significant infrastructure and economic challenges for researching agricultural pests impacting food security.

127

128 Material & Methods

129 Samples used in PCR primer efficacy tests

130 The DNA of whitefly cryptic species from Africa, Asia, South America, and Australia were individually
131 extracted using the Qiagen Blood and Tissue DNA extraction kit (Cat #. 69506). Extracted DNA of all samples to
132 be used for testing of PCR primer efficacies were individually eluted in 27.5 µL of EB. Whitefly species status
133 was confirmed by PCR and Sanger sequencing using the primer set wfly-COI-F1/R1 and/or wfly-COI-F2/R2
134 (Suppl. Material I). Primer development and Sanger sequencing followed the methods of Elfekih, Tay, et al.
135 (2018). Sanger sequencing was carried out at the BRF at JCSMR ANU, Canberra, ACT. Whitefly species used to
136 verify primer efficacies and their respective GenBank accession numbers are listed in Table 1.

137 **Table 1:** *Bemisia* species used to confirm PCR primer efficacies. With the exception of MN646951, all GenBank
138 accession numbers provided represent partial mtCOI sequences of *Bemisia* species tested where 100%
139 sequence identity against publicly available sequences have been detected. **Note:** KY951449 and MK386668**:**
140 nucleotide position: 782-1438.

Species	GenBank	Field sample identifiers	Country/source	Reference
MED-ASL	MH357342	60132	Malawi (sweet potato) Lat -12.90243413, long 34.28411598	Vyskocilova et al. (2019)
MEAM1	KY951449		Peru	Tay et al. (2017)
B. afer	AF418673	60266	Malawi, on ornamental tree cassava in township.	Maruthi et al. (2004)
Bemisia sp01	MN646951	30350	Tanzania (Lantana weed) Lat -2.742276, Long 33.021902	New *Bemisia* sp. (this study)
SSA1	JQ286457	40011	Tanzania (cassava) Lat -6.5221814, Long 38.924329	Mugerwa et al. (2012)
SSA2	AY057141		NRI laboratory culture	Legg et al. (2002)
SSA3	KM377923		NRI laboratory culture	Ghosh, Gowda, & Bouvaine

				(GenBank unpublished)
NW1	MK386668		Barbados	Kunz et al. (2019)
Asia II_1	MF497064		Bangladesh	Acharya & Kee (GenBank unpublished)
Asia1	MF497040		Bangladesh	Acharya & Kee (GenBank unpublished)
Lipaleyrodes (*Bemisia*)	MN056066		Australia	Kunz et al. (2019)

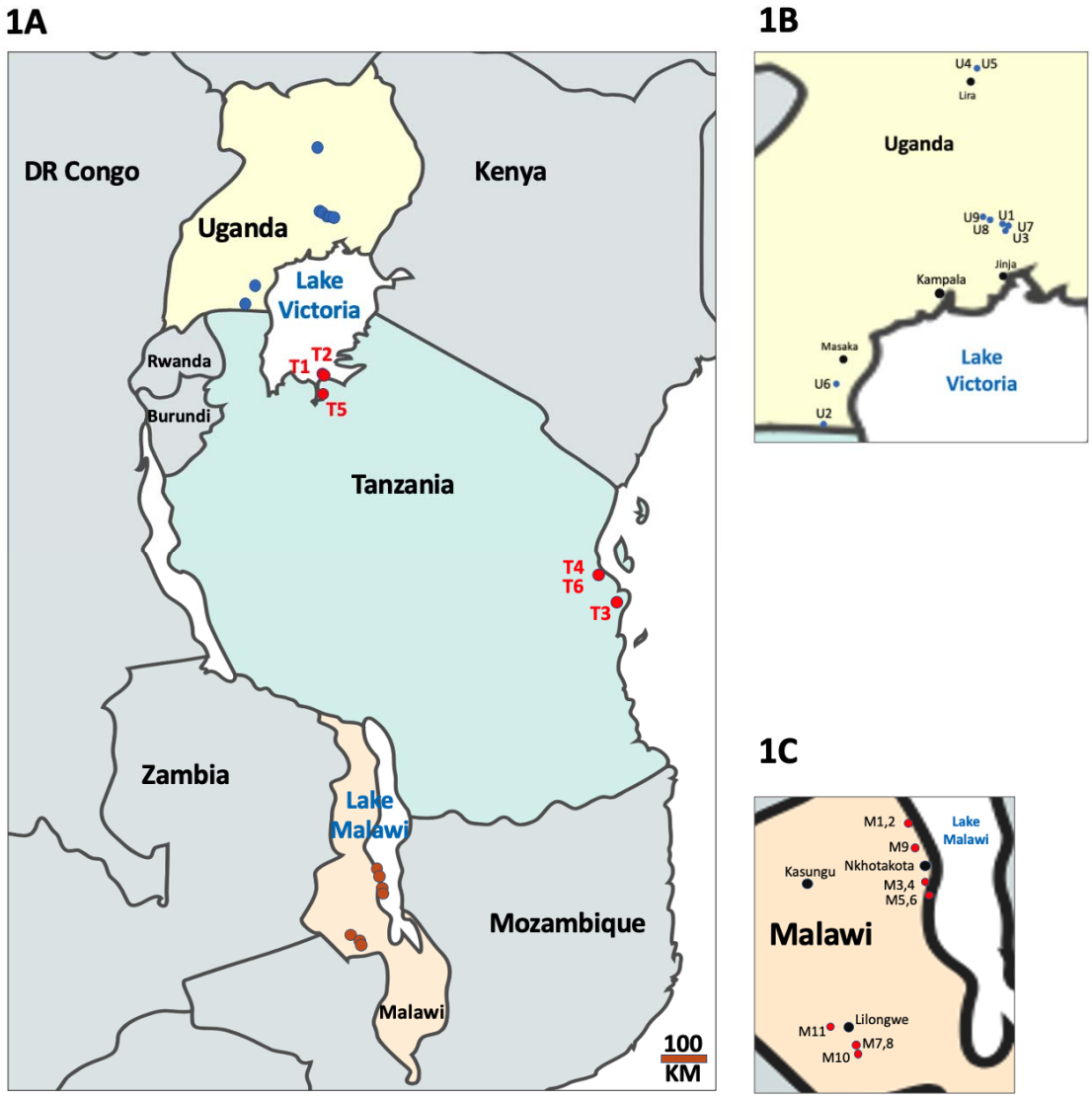
To ascertain PCR efficacies of the redesigned metabarcoding primers, various hypothetical field scenarios were simulated. These scenarios involved pooling of different genomic DNA (gDNA) that belonged to 11 separately extracted individuals of *Bemisia* species (Table 2) with predetermined species status based on the standard partial mtCOI dataset (Kunz, Tay, Elfekih, et al., 2019). These hypothetical scenarios included: (I) a two-species co-occurrence scenario, in which there is a dominant *Bemisia* whitefly species (e.g., SSA1) and a minor species present in different predetermined proportions (i.e., Table 2: Mixed-Ia (5%), Mixed-Ib (2.5%), and Mixed-Ic (1%)), and (II-IV) scenarios assuming co-occurrence of multiple mixed *Bemisia* cryptic pest species at different proportions based on estimates of gDNA concentration. With the exception of the test scenario 'Mixed-III' and 'Mix-IV' which were each replicated twice, all remaining test scenarios were replicated three times to estimate error rates and ensure protocol robustness.

Field samples survey methods

Cassava fields in Uganda, Tanzania, and Malawi were surveyed for *Bemisia* cryptic species between 2015-2016 (Fig. 1). In total, we conducted three data collection trips across a two-year period (trip 1, 1/8/2015 - 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 cassava fields to survey and sample for nymph and adult *Bemisia* whiteflies (for sampling details see also Macfadyen et al., 2021). Briefly, up to *ca*. 50 adults were collected using an aspirator from the top leaves of cassava plants of known variety in four Malawi field sites (Table 2). Nymphs were collected from all sites by selecting leaves that were lower down on the plant (and so had visible nymphs in the second to forth instar age range on the underside of the leaf). The top lip of two centimetre diameter circular plastic vial (5ml volume) was used to cut small leaf discs that contained nymphs. These discs were placed in the vial filled with 99% ethanol. As a comparison, nymphs from other crops surrounding the focal cassava field were also collected. All ethanol-preserved samples were stored in -20˚C prior to being processed in the laboratory,

where nymphs were dislodged from leaves and placed in fresh Eppendorf (1.5mL) test tubes in batches of between 20 and 40 individuals per field (note that in some cases, fields sampled for *Bemisia* contained <20 individuals). For our research question we selected nymphs that had no obvious signs of parasitism by hymenopteran parasitoids in order to optimize the *B. tabaci* DNA in the sample. We randomly selected 26 field-collected *Bemisia* populations from Malawi, Uganda and Tanzania for analysis to demonstrate the efficacies of the HTS mtCOI amplicon approach under field conditions (Table 3).

Fig. 1: (A) *Bemisia* cryptic species sampling sites from Uganda, Tanzania and Malawi, with field site codes from Tanzania (i.e., T1 – T6) provided in (1A), Uganda (i.e., U1 – U9) in panel 1B, and Malawi (i.e., M1 – M11) in panel 1C. GPS co-ordinates of each sampling site are provided in Table 3. Maps of Africa produced using Mapchart <https://mapchart.net>.


1A
DR Congo
Kenya
Uganda
Lake Victoria
Rwanda
Burundi
T1
T2
T5
Tanzania
T4
T6
T3
Zambia
Lake Malawi
Mozambique
Malawi
100 KM
1B
U4
U5
Lira
Uganda
U9
U8
U1
U7
U3
Jinja
Kampala
Masaka
U6
U2
Lake Victoria
1C
M1,2
Lake Malawi
M9
Kasungu
Nkhotakota
M3,4
M5,6
Malawi
M11
Lilongwe
M7,8
M10

175

176

rtain primer efficacies and species detection sensitivity by high-throughput sequencing platform. Average ± standard deviation of detection sensitivity (i.e., detection of observed proportion of taxa) for each category is provided based on amplicon reads as generated. 'Others' include amplicons associated with NUMTs (e.g., MEAM2; Tay, Elfekih, Court, et al., 2017), where the MEAM1 species was used in the Mixed-Ic and Mixed-IV scenarios), 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: † Only the first 10-15 amplicon contigs were searched for sequence identity by nucleotide blast (blastn) (i.e., optimised for somewhat similar sequences) search against the non-redundant (nr) NCBI DNA database. Verification of sequence identity was accomplished using the sanitized nucleotide database of Kunz, Tay, Elfekih, et al. (2019) as subject sequences, and using the NCBI's Basic Local Alignment Search Tool (BLAST) "blastn suite-2sequences" option and using only top hits (i.e., those with 97-100% sequence identity). PCR primer amplification efficacies were ascertained on species representing significantly divergent evolutionary lineages (Kunz, Tay, Elfekih, et al., 2019) that included a novel *Bemisia* species (MN646951; From Tanzania field code: 30350, see Macfadyen et al., 2021), Lipaleyrodes (*Bemisia*) from Queensland, Australia (MN056066), *B. afer,* the African cassava whitefly species (SSA1, SSA2, SSA3), and species within the *B. tabaci* cryptic complex from diverse geographic origins (e.g., New World 1 (NW1); Mediterranean (MED), Middle East Asia-Minor 1 (MEAM1); Asia II-1, and Asia 1).

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 reads can be downloaded from GenBank (SRA xxxx01, xxxx02, xxxx03). Site codes are as shown in Fig. 1.

Country	Site ID (site code)	Life stage	Host Plants (variety)	lat	long	Species composition	Total reads[#]	Reads Identified (%)
Malawi	60109/60121 (M1)	N, n= 26	Cassava (Gomani)	-12.639611255	34.17544665	**SSA1**: 12.7, **Baf**: 12.8, **Enc**: Y, **NUMTs**: 0.4	43,652	43,383 (99.35%)
	60109/60121 (M2)	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 (M3)	N, n=22	Cassava (Gomani)	-13.03129506	34.29746802	**SSA1**: 21.7, **NUMTs**: 0.17	68,562	68,224 (99.51%)
	60013/60126 (M4)	A, n= 8	Cassava (Gomani)	-13.03129506	34.29746802	**Baf**: 7.84, **NUMTs**: 0.09	44,424	44,166 (99.42%)
	60127/60250 (M5)	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 (M6)	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 (M7)	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 (M8)	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 (M9)	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 (M10)	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 (M11)	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 (U1)	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 (U2)	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 (U3)	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 (U4)	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 (U5)	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 (U6)	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 (U7)	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 (U8)	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 (U9)	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 (T1)	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 (T2)	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 (T3)	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 (T4)	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 (T5)	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 (T6)	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%)

Note: To calculate the estimated detected individuals for each species per sampling site, the proportion of amplicons assigned to each putative species was first determined by mapping to the reference species sequence (outlined in Fig. 2) and then divided by the total Illumina raw output fastq (i.e., Total reads[#]) reads from the pooled population.

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. Extraction of pooled individuals was life-stage specific (i.e., extraction of pooled adults or pooled nymphs were carried out separately). 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. Traditionally, molecular diagnostics for *Bemisia tabaci* cryptic species have relied on the F-C1-J-2195 and the TL2-N-3014 primers (Crozier, Crozier, & Mackinlay, 1989; Frohlich, Torres-Jerez, Bedford, Markham, & Brown, 1999; Roehrdanz, 1993; Simon et al., 1994) that amplified approximately 750bp of the 3' COI terminal region. This primer set was developed based on Diptera, Lepidoptera, Coleoptera, and Hymenoptera but was not originally reported for Hemiptera (Roehrdanz, 1993; Simon et al., 1994). Recent evaluations of these primers have found poor efficacies when applied in *B. tabaci* cryptic species (e.g., Elfekih, Tay, et al., 2018; Mugerwa et al., 2018) that contributed to the mis-identification of pseudo species (Kunz, Tay, Elfekih, et al., 2019; Tay, Elfekih, Court, et al., 2017). The *B. tabaci* cryptic species includes >34 well-defined species with interspecific nucleotide distances that ranged from *ca*. 3% to >18% (Dinsdale et al., 2010; Kunz, Tay, Elfekih, et al., 2019; Lee et al., 2013). To ensure maximal PCR efficacies of our primer sets, we aligned complete COI genes as obtained via HTS platform, from a range of species that included SSA1, SSA2 (Kunz, Tay, Elfekih, et al., 2019); MED cryptic species (Rossitto De Marchi, Kinene, Mbora Wainaina, Krause-Sakate, & Boykin, 2018; Tay, Elfekih, Court, et al., 2017; Vyskocilova et al., 2019); Asian species (Tay, Elfekih, Court, Gordon, & De Barro, 2016); *Bemisia* 'JpL' (Tay, Elfekih, Polaszek, et al., 2017), Australian (AUS), Indian Ocean, and Middle East Asian Minor 1 (MEAM1) species (Tay, Elfekih, Court, et al., 2017), to ascertain most conserved regions for designing and developing our two sets of complementary primers (wfly-PCR-F1/R1 and wfly-PCR-F2/R2; Suppl. Material 1). Primers were developed using Oligo 7 (Molecular Biology Insights, Colorado Springs, CO, USA) with minimal primer duplex and hairpin structure formation, with 55-65˚C Tm, for expected amplicon size of 550-600bp, and with minimal false primer annealing sites.

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 the PCR amplification of each pooled sample, we used 3.5ng gDNA template in a 35µL PCR reaction volume and 30 PCR cycles (PCR conditions described in

28 Suppl. Material I). Amplicons from each of the pooled samples were ascertained for PCR amplification success
29 on 1.25% 1x TAE agarose gel stained with GelRed and visualised on a UV-transilluminator. Amplicons were then
30 purified using AMPure XP beads as instructed in the Illumina **16S Metagenomic Sequencing Library**
31 **Preparation** (Part # 15044223 Rev. B) protocol and resuspended in 25 µL Qiagen EB (instead of the
32 recommended 52.5µL). Purified pooled amplicons were individually quantified for DNA concentration using a
33 Qubit v2.0 fluorometer (Life Technologies) and Qubit HS dsDNA assay kit and indexed by amplifying 5.0 ng of
34 the purified amplicon in 50µL PCR volume over 11 PCR cycles (see Suppl. Material I). Indexed amplicons were
35 purified using AMPure XP beads as described above and concentration ascertained using Qubit v2.0. Equal
36 proportions of purified indexed amplicon (i.e., 100ng) from each pooled sample were mixed with 6x Loading
37 Dye and run on 1.25% low melt agarose at 75 V for 2 hours. The F1/R1 combined amplicons and the F2/R2
38 combined amplicons were excised and purified using the Zymoclean Gel DNA Recovery Kit (cat. # D4001,
39 D4002, D4007, D4008) following the manufacturer's protocol. Purified indexed amplicons were quantified for
40 double stranded DNA concentration using Qubit and fragment size was estimated using TapeStation (Agilent
41 Technologies). Amplicons were calculated for their nano-molarity and diluted to 4 nM (see Illumina 470-2016-
42 007-B) for HTS run on MiSeq sequencer using the Illumina MiSeq Reagent Kit Version 3 (600bp pair-ended). A
43 detailed protocol for the amplicon library preparation and related primer sequence information is provided as
44 Suppl. Material I.

45 **HTS amplicon data processing**

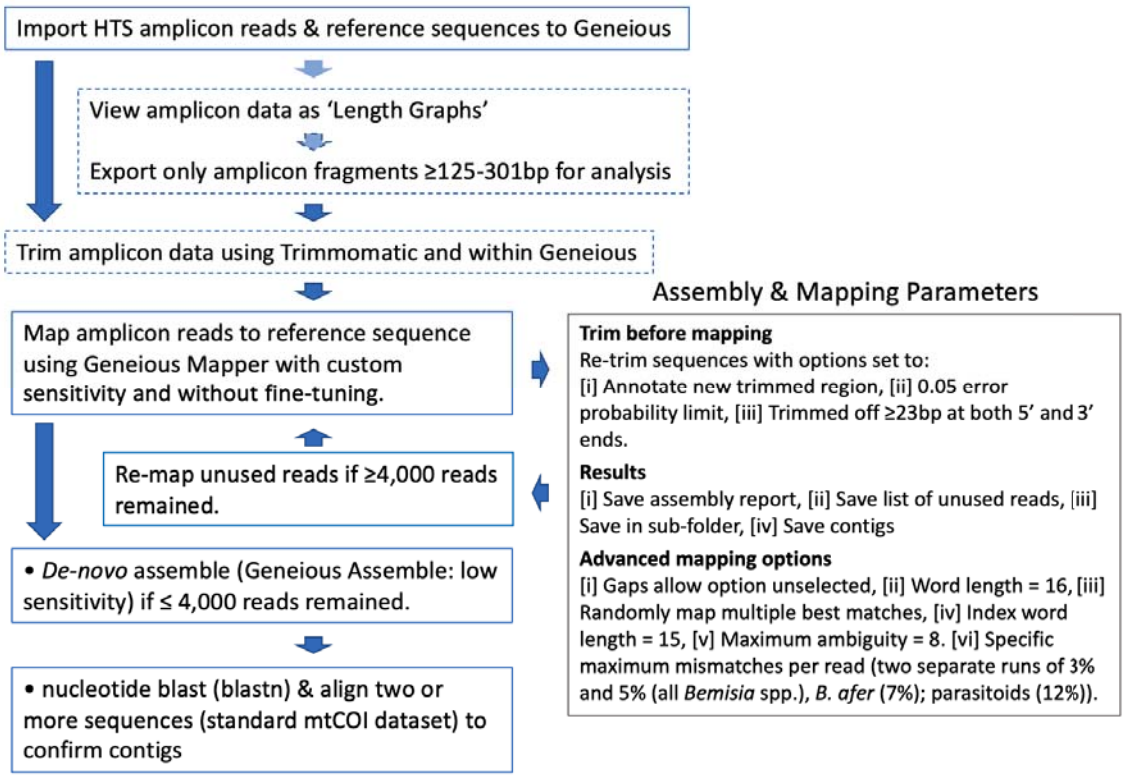
46 Fastq sequences representing individual populations post MiSeq runs were trimmed using
47 Trimmomatic version 0.36 (Bolger, Lohse, & Usadel, 2014) by quality (Leading:3; Trailing: 3;
48 Slidingwindow:4:15, minlength: 125bp) and imported into Geneious v11.1.5 as pair-ended reads of the two
49 metabarcoding primer sets for simultaneous mapping to reference sequences. Amplicon sequence reads from
50 each population were mapped to curated barcode reference sequences of representative *Bemisia* species (see
51 below) using the Geneious Mapper option, and selecting to trim amplicon reads with options set to 'Annotate
52 new trimmed region, error probability limit set to 0.05, and trimmed off ≥23bp at both 5' and 3' ends.
53 Customed mapping options within Geneious Mapper function were: no gaps allowed, word length set to 16;
54 randomly map multiple best matches, index word length set at 15, maximum ambiguity option set to 8, and

55 with two separate runs of 3% and 5% maximum mismatches per read for all *B. tabaci* species; 7% for *B. afer,*
56 and 12% for parasitoids (Fig. 2). Unmapped reads after each round were used in mapping against other *Bemisia*
57 and parasitoid reference sequences. A Geneious *de-novo* assembly step with default setting was carried out
58 separately for each primer-pair, and only so when all reference whitefly and parasitoid sequences had been
59 mapped and when un-mapped reads were typically low (i.e., ≤4,000 pair-ended reads). For the *de-novo*
60 assembly, the top 10 most common contigs were determined for their sequence identity using nucleotide blast
61 (blastn) within NCBI database. A schematic representation of the described workflow is provided in Fig. 2.

62

63 **Fig. 2:** Workflow for analysis of MiSeq generated high-throughput sequencing of mtCOI amplicon reads for
64 estimating proportions of *Bemisia* species in the African cassava cultivation landscape, including assessment of
65 hymenopteran parasitoid species. Optional steps are indicated in dashed box depending on whether amplicon
66 libraries were gel-purified (refer to Suppl. Material I) to remove small fragments. Quality of amplicons can also
67 be improved if desired, by using an Illumina sequence trimming program such as Trimmomatic (Bolger et al.,
68 2014). The *de-novo* assembly step enables novel *Trialeurodes-,* parasitoid- and bacterial-related sequences, as
69 well as NUMTs, to be identified. NUMTs were identified through detection of INDELs, premature stop codons,
70 and presence of unexpected amino acid residues at highly conserved COI regions following the methods of
71 Kunz, Tay, Elfekih, et al. (2019).

72



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. Our reference mtCOI sequences were from the following *Bemisia* species: SSA1 (GenBank JQ286457, nt 59-715), SSA2 (GenBank AY057141, nt 29-685), SSA3 (GenBank KM377923, nt 20-676), MEAM1 (GenBank KY951449, nt 782-1438), IO (GenBank KY951448, nt 782-1438), *B. afer* (GenBank AF418673, nt 59-715), Uganda1 (GenBank KX570857, nt 34-690), 'African' MEAM2 (GenBank KX570778, nt 34-690), MED (GenBank KY951447, nt 782-1438), SSA6 (GenBank KX570852, nt 34-690), SSA9 (GenBank KX570856, nt 34-690), SSA10 (GenBank KX570804, nt 34-690), SSA13 (GenBank KX570833, nt 34-690) and novel sub-Saharan *Bemisia* species (i.e., MN646951; MN646952) detected by Sanger sequencing from field-surveyed populations (Macfadyen et al., 2021; this study). We confirmed the authenticity of these reference sequences as free from NUMTs following the methods of Kunz, Tay, Elfekih, et al. (2019).

Parasitoids reference sequences

87 *Eretmocerus* and *Encarsia* parasitoid mtCOI reference sequences from the sub-Saharan region
88 surrounding our field survey sites were built up (MN646919, MN646949) during the course of the analysis
89 workflow (see Fig. 2). Due to the anticipated significant whitefly hymenopteran parasitoid species diversity in
90 the sub-Saharan region, we arbitrarily determine a cut-off value of 6% within the Geneious 'Map to Reference'
91 function.

92 **Assessment of species delimitation using a stagged over-lapping amplicon approach**

93 To assess efficacies of our two-primer set approach for metabarcoding of cryptic *Bemisia* and related
94 whitefly species, we downloaded the clean *B. tabaci* mtCOI database of Kunz, Tay, Elfekih, et al. (2019) and
95 performed phylogenetic analyses based on 315bp aligned nucleotides at 50bp sliding window sizes for primer
96 pairs 'wfly-PCR-F1/R1' (i.e., 3' mtCOI barcode nucleotide position (nt) 1-315; nt 51-365; nt 101-415; nt 151-
97 465), and 'wfly-PCR-F2/R2' (i.e., nt 140-454; nt 190-504; nt 240-554; nt 290-604; nt 340-654), as well as for the
98 full length of wfly-PCR-F1/R1 amplicon (473bp) and the wfly-PCR-F2/R2 amplicon (518bp), and compared that
99 with the species delimitating powers based on uncorrected pair-wise nucleotide similarity (i.e., *p*-dist) distance
100 method of the full 657bp 3' mtCOI *B. tabaci* barcoding gene. Sequence alignment using MAFFT Alignment
101 (selecting 'Auto' option for the Algorithm setting, 200PAM / K=2 scoring matrix, and gap open penalty and
102 offset value set at 1.53 and 0.123, respectively) (Katoh & Standley, 2013) and trimming were carried out in
103 Geneious v11.1.5. Confidence of clustering of identified species was estimated using Ultrafast bootstrap as
104 implemented in IQ-Tree, and followed the methods as described in Kunz, Tay, Elfekih, et al. (2019). Briefly,
105 aligned and trimmed sequences were imported as fasta formatted file and imported to the IQ-Tree website
106 <http://iqtree.cibiv.univie.ac.at> (Trifinopoulos, Nguyen, von Haeseler, & Minh, 2016). The 'Automatic'
107 substitution model selection option was selected and we specified 1,000 ultrafast bootstrap approximation
108 (Minh, Nguyen, & von Haeseler, 2013) with 1,000 maximum iterations and 0.99 as the minimum correlation
109 coefficient. Visualisation and assessment of correct clustering of species were carried out using FigTree v1.4.4
110 (2006-2018 Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh).

111 **Results**

112 **Primer efficacies:** The primers wfly-COI-F1/R1 and/or wfly-COI-F2/R2 were shown to be highly efficient in
113 successful PCR amplification of diverse representative *Bemisia* species that included Australian species

114 (Lipaleyrodes (*Bemisia*)), Asian, New World (NW), Mediterranean (MED), Middle East Asian Minor (MEAM1),
115 sub-Saharan African (SSA) species, and 'non-*tabaci*' species (i.e., *B. afer*). Sanger sequencing of PCR amplicon
116 from test *Bemisia* cryptic species (Table 1) partial mtCOI confirmed species identity when compared against
117 both the updated standard *B. tabaci* mtCOI reference dataset and the uncorrected pairwise nucleotide
118 distances (*p*-dist; see Kunz, Tay, Elfekih, et al., 2019). While Sanger sequencing confirmed sequence integrity
119 such as through conservation of amino acid residues against published *B. tabaci* cryptic species HTS-derived full
120 mtDNA genomes (e.g., Kunz, Tay, Court, et al., 2019; Tay, Elfekih, et al., 2016; Tay, Elfekih, Court, et al., 2017;
121 Tay, Elfekih, Polaszek, et al., 2017; Vyskočilová et al., 2018), results from HTS amplicon analyses readily
122 detected NUMTs such as those similar to 'MEAM2' (Delatte et al., 2005; Tay, Elfekih, Court, et al., 2017), as well
123 as partial mtCOI genes with INDELs. Detection of co-amplified pseudogenes/NUMTs demonstrated the
124 complexity of the *Bemisia* genome (Chen et al., 2016; Xie et al., 2017; Xie et al., 2018). HTS-amplicon analyses
125 also shed light on the unexpected co-amplification of non-target whitefly genomic regions, as well as fungal,
126 bacterial symbionts, and parasitoids genomic regions, and highlighted the complexity of trophic interactions in
127 *Bemisia* whiteflies (e.g., Rao et al., 2018; Shamimuzzaman et al., 2019; Tay, Elfekih, Polaszek, et al., 2017).

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

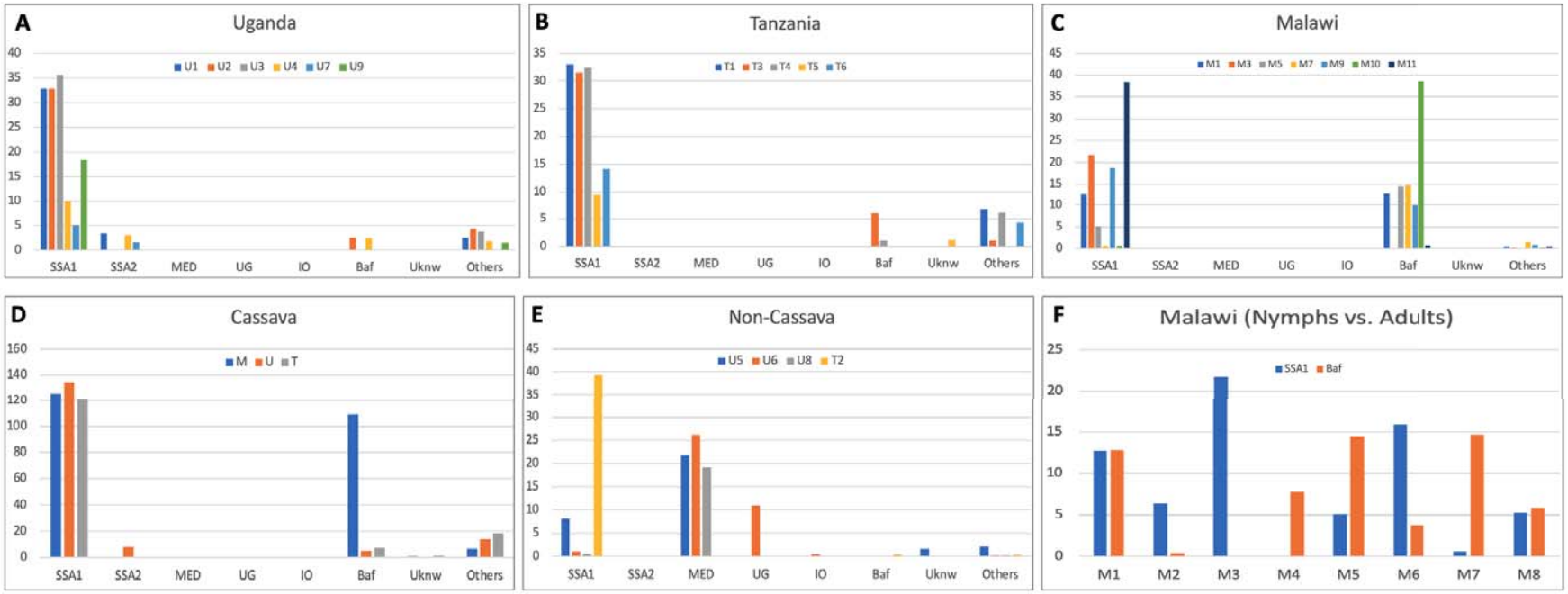
139 **Field application of HTS-amplicon method:** Applying the HTS multi-primer metabarcoding protocol to field-
140 collected samples from Malawi, Tanzania, and Uganda showed that in the African cassava fields, differences in

141 *Bemisia* species compositions existed (Kalyebi et al., 2018; Macfadyen et al., 2018; Macfadyen et al., 2021). A
142 total of eight whitefly species (i.e., SSA1, SSA2, IO, MED, *B. afer,* Uganda 1, two novel *Bemisia* species), and a
143 novel *Trialeurodes* species, were detected in different field sites (Table 3, Fig. 3). Multiple novel mtCOI
144 haplotypes for *Eretmocerus* species (MN64915-MN64927) and *Encarsia* species (MN646929-MN646950) were
145 also detected (Table 3). Parasitoid sequences were only ever detected in nymphal samples, and this was
146 despite targeted exclusion of parasitised nymphal samples that could be visually identified prior to genomic
147 DNA extraction. Detection of parasitoids is therefore likely the result of early stages of parasitism where visual
148 identification would be most challenging. While our findings suggested high parasitism rates in the agricultural
149 landscapes in Uganda, Tanzania, and Malawi, we nevertheless advise caution in interpreting the *Bemisia*
150 parasitism rates, parasitoid species, and genetic diversity in these natural agricultural settings, as our primary
151 experimental design was to develop a metabarcoding approach for surveys of the hemipteran *Bemisia* cryptic
152 pest species only.

153 The sub-Saharan African 1 (SSA1) *Bemisia* species was identified as the dominant species in cassava
154 fields from Uganda and Tanzania (Figs. 3A, 3B), while in Malawi both SSA1 and *B. afer* were present in
155 approximately similar proportions (Fig. 3C). Overall, the SSA1 species was most prevalent in cassava landscape
156 (Fig. 3D). We also detected the SSA1 species on non-cassava crop planted within a cassava field ('T2' in Fig. 3E;
157 Table 3). The *B. tabaci* MED cryptic species complex was present as the dominant species in Uganda from three
158 non-cassava host crop sampling sites (Fig. 3E; Table 3). Sweet potato crops also hosted other *Bemisia* species
159 including Uganda 1 ('UG') and Indian Ocean ('IO'), although both UG and IO species were not detected on
160 sweet potato within cassava fields in Tanzania ('T2'), and could suggest potential behavioural differences
161 between SSA1 and UG/MED species (Fig. 3E) impacting on species-specific abundances. In Malawi, *B. afer* and
162 SSA1 species dominated different field sites (Table 3, Figs. 3C, 3F). Their detection showed significant variability
163 in species composition within the same field sites depending on the life-stage sampled (i.e., nymphal vs. adults;
164 Fig. 3F). Field sample analyses also detected NUMTs and bacterial/fungal origins in the African populations,
165 similar to our hypothetical scenarios (Table 2), thereby demonstrating consistencies between our HTS
166 workflow and primer-efficacy assessments.

167 **Assessment of species delimitation efficacies between amplicon sizes:** Short amplicons of between 130bp –
168 319bp (e.g., Brandon-Mong et al., 2015; Leray et al., 2013) but also longer amplicon lengths (e.g., 421bp,
169 Hajibabaei, Porter, Wright, & Rudar, 2019) have been proposed for COI metabarcoding. Our assessment of
170 species delimitation powers by short (i.e., 315bp) versus full amplicon lengths based on distance methods (i.e.,
171 p-dist) of our wfly-PCR-F1/R1 (473bp) and wfly-PCR-F2/R2 (518bp) primer sets, as well as their combined
172 657bp barcode gene length for the *B. tabaci* cryptic species complex showed that metabarcoding involving the
173 full 657bp was needed to fully delimit between *Bemisia* species (Fig. 4). Species detection and delimitation
174 involving the various shorter 315bp nucleotide lengths along the 657bp barcoding gene region often failed to
175 differentiate certain cryptic species complexes such as those within the Mediterranean (MED), New World
176 (NW), sub-Saharan Africa (SSA), and Australia (AUS) clades (Fig. 4). Interestingly, other invasive species such as
177 the SSA2 (e.g., Hajibabaei et al., 2019) and the MEAM1 (De Barro et al., 2011; Elfekih, Etter, et al., 2018), as
178 well as the Asia II-1 species that has recently emerged as important vector species in the transmission of the Sri
179 Lankan Cassava Mosaic Virus (SLCMV) in South East Asia (Chi et al., 2020) could be readily identified by any of
180 the 315bp COI region.

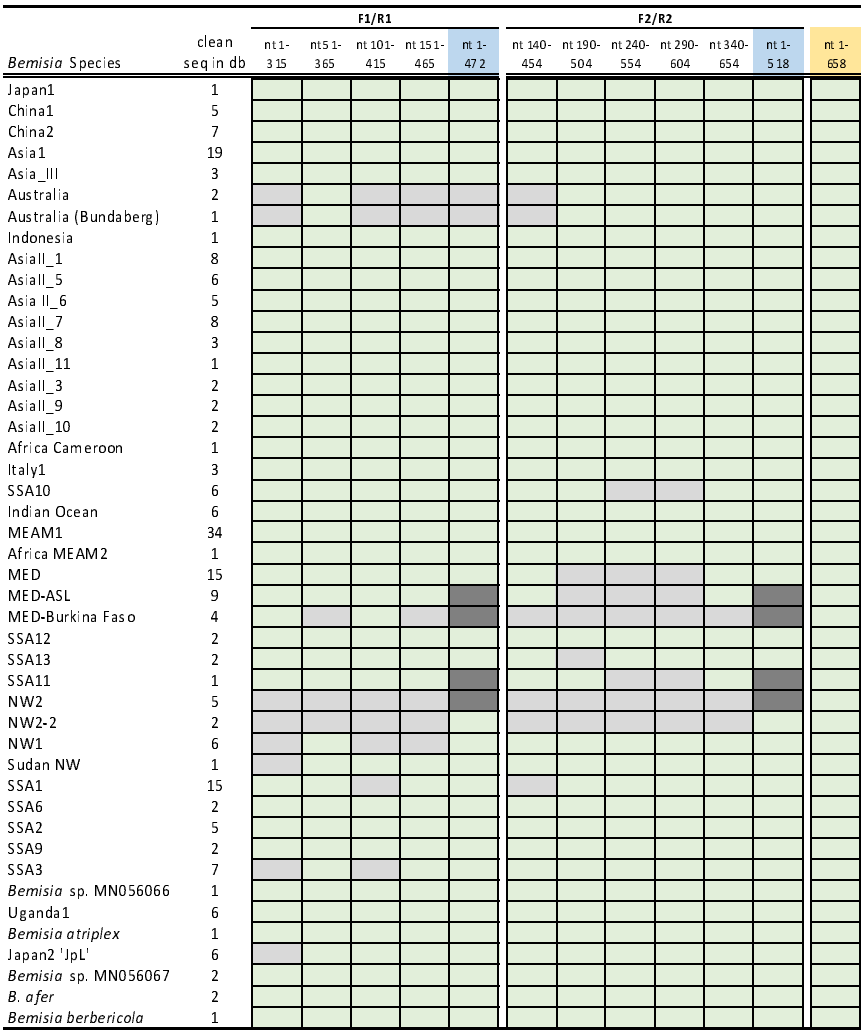
181 **Fig. 3:** Summary results of *Bemisia tabaci* cryptic species from Uganda (panel A), Tanzania (panel B), and Malawi (panel C). Site codes are as provided in Table 3.
182 *Bemisia* species are sub-Sharan Africa 1 (SSA1), sub-Saharan Africa 2 (SSA2), Mediterranean (MED), Uganda 1 (UG), Indian Ocean (IO), *B. afer* (Baf), and unknown
183 ('Uknw') species from *Bemisia* and *Trialeurodes* genera. Low detection rates of COI-related pseudogene, parasitoid (*Eretmocerus* and *Encarsia* genera), and
184 bacterial/fungal sequences are grouped in the 'others' category. Whitefly species compositions are also compared between cassava and non-cassava host plants
185 (panels D and E, respectively), and between nymphal (M1, M3, M5, M7) and adult (M2, M4, M6, M8) life stages from four cassava sites in Malawi (panel F).



186

187

Fig. 4: Summaries of efficacies of species delimitation powers based on the *p*-dist method across the 657bp 3' mtCOI barcoding gene region in the cryptic *Bemisia tabaci* and non-*tabaci* species complex. The metabarcoding primer pair 'wfly-PCR-F1/R1' (472bp) was divided into four overlapping regions of 315bp across a 50bp sliding window size, and the 518bp amplicon generated by the 'wfly-PCR-F2/R2' primer pair was divided into five overlapping regions of 315bp. We also assessed full amplicon lengths from both primer pairs (i.e., F1/R1: 472bp; F2/R2: 518bp, blue coloured cells). Cryptic species that could not be defined were represented by grey coloured cells, and sequences that were successfully defined into their respective species clades were shown by green cells. Dark grey cells indicated failed species delimitation by both primer pairs and full species delimitation will require the complete 657bp barcoding gene. The number of clean mtCOI sequences in the database (db) of Kunz, Tay, Elfekih, et al. (2019) for each species analysed are indicated (total: 224 sequences).



199 **Discussion**

200 This study successfully demonstrated the applicability, transferability and powers of the multi-primer
201 metabarcoding approach to investigate species compositions within an agroecological context, using the highly
202 challenging *Bemisia* whitefly cryptic pest species as a study model system. These complementary sets of
203 primers developed to amplify across the large genetic distances among the *B. tabaci* species complex also
204 detected non-*Bemisia* hemipteran whitefly-nymphs, while showing potential to also simultaneously survey for
205 associated parasitoid species diversity and parasitism rates, thereby further demonstrating the wide
206 applicability and contribution to our understanding of trophic interactions between pests, beneficial insects
207 and host plant usage across highly heterogeneous agricultural landscapes. Our multi-primer metabarcoding
208 approach that involved high-volume sampled biological material demonstrated host plant preference and
209 species composition differences in African *Bemisia* species (Table 3, Fig. 3), and that ascertaining a *Bemisia*
210 species' taxonomic status will require the standard length (i.e., ≥657bp) mtCOI barcoding 3' gene region (Suppl.
211 Table 1; Fig. 4), in addition to considering also developmental/life stages (Fig. 3F), and the species' plant virus
212 vectoring capabilities (e.g., Chi et al., 2020). We also outlined the development of an easy-to-use tailored
213 bioinformatic analysis workflow, and provided the *Bemisia* research communities with two sets of highly
214 efficient PCR primers that have the capability to amplify the alternative but widely used 3' mtCOI *Bemisia*
215 barcoding region. Our primer sets when used in metabarcoding of *Bemisia* nymphs, also detect various
216 Aphelinid hymenopteran parasitoid species within the *Eretmocerus* and *Encarsia* genera. Unexpected co-
217 amplification of various fungal and bacterial homologous gene regions, albeit at low frequencies, was also
218 detected.

219 DNA metabarcoding approaches have been reported for integrative taxonomy (e.g., Cruaud, Rasplus,
220 Rodriguez, & Cruaud, 2017; Sigut et al., 2017), their potential for invasive species surveys reviewed (Piper et al.,
221 2019), for understanding relationships between species abundances and associated ecological factors (e.g.,
222 Macfadyen et al., 2021), and impact from aquacultural activities demonstrated (He et al., 2021). Adoption of
223 such metabarcoding approaches for landscape ecological investigations under field conditions between insect
224 pest hosts and beneficial insects may also be possible (e.g., Macfadyen et al., 2021), although appropriate
225 experimental designs, primer improvement and modifications to analysis workflow including building of

226 relevant barcoding database will be needed to increase species survey and detection sensitivity. For example,
227 DNA barcoding primers for arthropods including the Hymenoptera (e.g., Folmer, Black, Hoeh, Lutz, &
228 Vrijenhoek, 1994; Hebert, Penton, Burns, Janzen, & Hallwachs, 2004; see also Woodcock et al., 2013) could be
229 included as internal checks to compare frequencies of parasitism and species diversity, with concurrent
230 analysis following the HTS pipeline as outlined (Fig. 2). Studies that aimed to understand host-parasitoid
231 interactions involving Aphalinid parasitoids faces other challenge such as high species diversity in the
232 Aphelinidae (e.g., Gebiola et al., 2017; Hernández - Suárez et al., 2003), while different reproductive modes
233 (sexual vs. asexual vs. Rickettsiales-induced sex ratio biases) in these haplodiploid parasitoids further
234 complicate our understanding and interpretation of such biological interactions. Efficacies of these two
235 complementary primer sets described in this study on hymenopteran parasitoids requires further assessment,
236 as primer choice are known to affect species diversity recovery rates (e.g., Hajibabaei et al., 2019) and is
237 currently beyond our research aims.

238 Primer efficacies is a significant issue for *Bemisia* molecular diagnostics (e.g., Elfekih, Tay, et al., 2018;
239 Mugerwa et al., 2018; Shatters et al., 2009), and have contributed to mis-identification of pseudo-species (e.g.,
240 MEAM-K, Roopa et al., 2015; MEAM2, Delatte et al., 2005; Karut, Kaydan, Tok, Doker, & Kazak, 2015; Ueda,
241 Kitamura, Kijima, Honda, & Kanmiya, 2009; SSA4, Berry et al., 2004) as demonstrated through genomic
242 approaches (Elfekih et al., 2019; Kunz, Tay, Elfekih, et al., 2019; Tay, Elfekih, Court, et al., 2017; Vyskočilová et
243 al., 2018). Our two primer sets while successful in test scenarios in amplifying diverse *Bemisia* cryptic species
244 (Table 2), by themselves, have nevertheless failed to resolve species status within various cryptic species
245 complexes (Fig. 4) due to insufficient DNA length. In a previous study by Wang et al. (2018), it was proposed
246 that cost-effective NGS barcoding solutions were available albeit with shorter (e.g., 313bp) partial mtCOI
247 sequences, however this would result in poor study outcome due to the biological and taxonomical complexity
248 associated with the *B. tabaci* species complex (see Fig. 4). When used together however, was able to
249 differentiate all but the closely related MED-ASL/MED-Burkina Faso species complex, SSA11, and also the two
250 closely related NW2/NW2-2 species complex. Differentiating between certain cryptic *Bemisia* species may
251 benefit from using alternative DNA markers and gene regions. For example, Vyskočilová et al. (2018) showed
252 that species delimitation of the MED-ASL from other MED cryptic species was best achieved using the standard

253 (i.e., 5') barcoding gene region or via full mitogenomes. Furthermore, delimiting the true identity of the SSA1
254 species complex in our study sites will likely require genome-wide single nucleotide polymorphic markers
255 (Elfekih et al., 2019). For the putative NW2/NW2-2 species complex identified by Kunz, Tay, Elfekih, et al.
256 (2019), it remained to be seen if the 5' COI gene region or if genome-wide single nucleotide polymorphic
257 makers would prove to be the more effective diagnostic marker system.

258 Test scenarios involving evolutionarily divergent whitefly species (Tables 1 and 2) and field data on
259 diversity of whitefly species detected (e.g., included novel *Bemisia* and *Trialeurodes* species; Table 3) suggested
260 that overall non-detection rates of species could be potentially low. Our primers further demonstrated
261 unexpected detection of novel *Encarsia* and *Eretmocerus* parasitoid species in the African cassava and non-
262 cassava fields, however the true genetic and species diversity of these minute parasitoids will require
263 independent assessment for confirmation, such as to include the standard barcode COI gene region using
264 standard primers (e.g., C_LepFolF/C_LepFolR; Woodcock et al., 2013) as internal control. Nevertheless,
265 understanding of trophic interactions between crop hosts, pest/beneficial insects and their potential biological
266 control agents represents a molecular ecological network research area that could yield significant insights via
267 metabarcoding approach (e.g., Bansch et al., 2020; Evans, Kitson, Lunt, Straw, & Pocock, 2016; Sow et al.,
268 2019).

269 Determining which insect represents the major species within diverse agricultural landscape settings
270 requires adequate population-wide sampling (Kalyebi et al., 2021; Kalyebi et al., 2018; Macfadyen et al., 2021).
271 Conclusions based on few individuals sampled at a large number of sites are inherently problematic; there is a
272 need, therefore, for the development of techniques to scale up data collection and processing of samples, and
273 to develop associated bioinformatic workflows and analytical pipelines to provide robust findings at economic
274 scales. For the cassava cultivation landscape in sub-Saharan Africa, determining the dominant pest whitefly
275 cryptic species represents a significant challenge due to difficulties with species identification without the help
276 of molecular DNA markers (De Barro et al., 2011; Kunz, Tay, Elfekih, et al., 2019). Furthermore, the often-high
277 population sizes associated with individual fields, and poor PCR primer efficacies (Elfekih, Tay, et al., 2018;
278 Mugerwa et al., 2018; Shatters et al., 2009) regularly resulted in suboptimal PCR amplification, and the
279 difficulty encountered in sequence quality control (Kunz, Tay, Elfekih, et al., 2019; Tay, Elfekih, Court, et al.,

280 2017; Vyskočilová et al., 2018), have been persistent problems in *Bemisia* molecular systematics and
281 diagnostics.

282 In this study, we showed that by applying the DNA barcoding and the species 'genetic gaps' concepts
283 first demonstrated by Dinsdale et al. (2010; see also Kunz, Tay, Elfekih, et al., 2019), and with careful re-
284 designing of DNA markers for the HTS metabarcoding platform, it was possible to develop a highly efficient
285 method for large-scale, effective and economic sampling and identification of the cryptic whitefly species to
286 contribute to their management and control in challenging agricultural landscape settings (Kalyebi et al., 2018;
287 Macfadyen et al., 2018; Macfadyen et al., 2021). Our species survey findings showed that SSA2 and *B. afer*
288 were also present in the East Africa's cassava landscape in addition to the dominant SSA1 species, and
289 contrasted the reported absence of SSA2 based on sampling of low numbers of individuals in Ugandan cassava
290 landscape (Ally et al., 2019). Co-exitance of both SSA1 and SSA2 may have important implication to genome
291 evolution through introgression (Elfekih et al. 2021) and emergence of cassava disease epidemic (Legg,
292 Sseruwagi, et al., 2014; Patil & Fauquet, 2009). Significant *Bemisia* whitefly species diversity such as the Asia
293 species complex (e.g., see Kunz, Tay, Elfekih, et al., 2019) as well as invasive MED and MEAM1 species existed
294 across the Asian continent. Adopting our metabarcoding approach to delimit species status, composition,
295 prevalence, and their association with asymptomatic/diseased cassava plants could provide knowledge
296 necessary to assist with the development of management strategies for the recent SLCMV disease outbreaks in
297 the South East Asian region (Chi et al., 2020; Minato et al., 2019).

298 The importance of using standardised approach to sample sufficiently large quantities of *B. tabaci*
299 cryptic species (Sseruwagi, Sserubombwe, Legg, Ndunguru, & Thresh, 2004; Macfadyen et al., 2021) especially
300 across diverse agricultural landscape and between different life stages (e.g., nymph vs. adult) is also
301 demonstrated (e.g., Table 3, Fig. 3F), where we detected significant differences in species compositions (e.g.,
302 between *B. afer* and SSA1). The approach developed here is therefore useful for processing nymphal data that
303 may have strong links to host plants. Standardized sampling techniques for nymphs are less well-developed
304 than for adult *B. tabaci* on cassava (which are considered easier to count and collect). In reality, the focus of
305 the research question should drive the sampling approach selected (be that nymphs, parasitized nymphs, or
306 adults); however, in all cases scaling up the numbers of individuals processed is critical. Another important

307 distinction of specific sampling of whiteflies from different host crops was also demonstrated in cassava versus
308 non-cassava host crops (Table 3, Figs. D, E), where differences in species compositions, especially between *B.*
309 *afer* and SSA1 *cf.* MED, Ug, and IO, respectively, were evident. Ecological and biological factors therefore
310 underpinning *Bemisia* species compositions in Africa agricultural landscapes. Similarly, ecological factors and
311 agricultural practices will likely impact on pest species compositions in other agro-ecological landscape systems
312 and can be assessed by adopting the multi-primer metabarcoding method outlined here.

313 While our method enables efficient ascertainment of species composition at the landscape level, this
314 method should not be used to determine individual maternal lineages of mtDNA haplotypes due to the high
315 sensitivity of detecting random PCR-introduced SNPs in individual amplicon molecules. At the landscape survey
316 level, ascertaining specific maternal lineages does not represent a criterion needed to enable species identity
317 to be determined, provided that sequence quality was adequately managed to avoid misidentification of
318 species status due to mtDNA pseudogenes. Elfekih et al. (2021) have shown that NUMTs and pseudogenes
319 could lead to significant misinterpretation of *Bemisia* population structure in the African cassava landscape
320 regardless of whether partial mtCOI or genome-wide SNPs were used (see e.g., Wosula, Chen, Fei, & Legg,
321 2017). A significant limitation on our approach is the quality of the library used to interrogate the sequences
322 generated. Sanger sequencing-generated partial mtCOI gene sequences in the public database (e.g., NCBI; see
323 Vyskočilová et al. (2018) regarding MED-related pseudogenes/NUMTs; see also Kunz, Tay, Elfekih, et al. (2019)
324 for reanalysis of *B. tabaci* DNA database (Boykin, Savill, & De Barro, 2017)) are gradually being recognised for
325 containing high numbers of problematic sequences (see e.g., Nacer & do Amaral, 2017; Song, Buhay, Whiting,
326 & Crandall, 2008; Triant & Dewoody, 2007). Such readily available public DNA global databases will need to be
327 critically reviewed and assessed, given that a large scientific community relies on these sequences to assist with
328 species identification.

329 A further complicating factor in molecular diagnostics of *Bemisia* species is the significant challenge
330 associated with their nomenclature, cryptic taxonomy (Boykin et al., 2018; De Barro et al., 2011), and the
331 preferred partial COI gene region widely adopted by the *Bemisia* tabaci research community (i.e., 3' COI region)
332 against the standard barcode 5' COI region. For example, against the 34 *B. tabaci* cryptic species and five non-
333 *tabaci* species presented by Kunz, Tay, Elfekih, et al. (2019) based on sanitized dataset, Barcode of Life data

334 system v4 (BOLDSYSTEMS) did not identify 30 of these, while eight cryptic species were all identified as *B.*
335 *tabaci,* and only one was correctly identified as *B. afer* (Suppl. Table 1). Furthermore, within the BOLDSYSTEMS
336 database, the invasive MEAM1 was identified as '*B. tabaci*', while the Mediterranean (MED) species, shown as
337 the real *B. tabaci* species as originally described by Gennadius in 1889 (Tay et al., 2012), returned as 'no species
338 level match'. While BOLDSYSTEMS database and the NCBI database are linked, inherent problems associated
339 with reporting cryptic species remained (e.g., unable to provide cryptic species identity in NCBI sequence
340 submission process) and will continue to impact on research outcomes especially as metabarcoding becomes
341 more widely and rapidly adopted. Reporting of full COI gene such as through assemblies of draft mitogenomes
342 in the *Bemisia* whitefly species complex (e.g., Kunz, Tay, Court, et al., 2019; Tay, Elfekih, Court, et al., 2017; Tay,
343 Elfekih, Polaszek, et al., 2017; Thao & Baumann, 2004; Vyskočilová et al., 2018; Wang et al., 2016) can
344 consolidate the different databases for *B. tabaci* cryptic species. Linking of other non-standard gene regions
345 (e.g., mitochondrial DNA 12S rRNA and ND2 gene regions) to the standard 5' COI gene region will also further
346 resolve our understanding of species diversity from diverse ecosystems (e.g., Miya, Gotoh, & Sado, 2020;
347 Stefanni et al., 2018; Weitemier et al., 2021).

348 Our metabarcoding platform can be adapted to understand other pest species complexes in
349 agricultural systems and may be especially effective for invasive insect pests with small body size and where
350 high density of population sizes can build up rapidly, such as in invasive whiteflies, thrips, mites, and aphids.
351 We propose, therefore, to apply this approach to: (1) greatly increase number of individuals that can be
352 sampled, processed and characterised simultaneously for their widely applied proposed 5' COI (Hebert et al.,
353 2003) or alternative 'DNA barcoding' genes and gene regions (e.g., the mtCOI 3' region), and the earlier
354 (Woese, Kandler, & Wheelis, 1990) proposed RNA genes and related regions (i.e., ITS rRNA region for fungi,
355 e.g., Xu (2016); Microsporidia, e.g., Ghosh & Weiss, 2009; Klee, Tay, & Paxton, 2006; O'Mahony, Tay, & Paxton,
356 2007; Tay, O'Mahony, & Paxton, 2005; Velasquez et al., 1996; 16S rRNA, e.g., Shelomi & Chen, 2020; 18S rRNA,
357 e.g., Popovic et al., 2018; 28S rRNA, e.g., Pawlowski et al., 2012); (2) estimate individual species proportion
358 across large agricultural landscapes, and (3) in nymphal/larval samples, identify parasitoid species that target
359 the agricultural pests of interest. Contrasting the majority of metabarcoding studies published to-date, our
360 approach utilises significantly longer partial COI gene sequence through two sets of complementary primer sets

361 (i.e., *ca*. 650bp *cf*. 130-421bp; Brandon-Mong et al., 2015; Hajibabaei et al., 2019; Miya et al., 2020; Stefanni et
362 al., 2018) which was important for defining cryptic *Bemisia* species status. The use of this multi-primer
363 metabarcoding approach to generate longer sequence length and targeting the 5' COI barcoding gene region
364 could enable greater species identification with the support of the BOLDSYSTEMS.

365 While the high-throughput amplicon sequencing method in a species composition study in
366 agroecological landscapes represents a relatively new field of research, the use of this method to identify genes
367 of biosecurity importance has been demonstrated for understanding the spread of pyrethroid resistance genes
368 in the red-legged earth mites *Halotydeus destructor* (Edwards et al., 2018). The method described by Edwards
369 et al. (2018) can be incorporated into our high-throughput amplicon sequencing platform for rapid species
370 composition identification and insecticide resistance allele frequency estimates. This would enable a genomic
371 approach that would simultaneously allow the identification of species composition at landscape scales, the
372 survey of insecticide resistance profiles (e.g., Hopkinson et al., 2019), and assessment of the implication of
373 chemical control strategies on diversity and density of beneficial insects such as parasitoids. Such study would
374 be especially informative when temporal scales are included within the study designs (e.g., Kalyebi et al., 2021;
375 Macfadyen et al., 2021).

376 Developmental stages sampled do impact *Bemisia* species compositions and the conclusions drawn
377 from a survey, as demonstrated in the Malawi cassava sampling sites for adult and nymphal *Bemisia*
378 populations. The extent of such variability in target species composition between immature and adult life
379 stages for many other biological systems can also benefit from the HTS metabarcoding method described here.
380 In this study, *Bemisia* nymphs that visibly showed signs of being parasitised by hymenopteran parasitoids were
381 first excluded prior to genomic processing. Despite this, *Encarsia* and *Eretmocerus* parasitoid mtCOI sequences
382 were detected, thereby highlighting the importance of biological control of *Bemisia* pest species by parasitoids
383 in the sub-Saharan African agricultural system. While we have not optimised this workflow to address host-
384 parasitoid interaction research questions, we have demonstrated the power of HTS genomic approaches to
385 document and facilitate an understanding of these biological interactions. Improved understanding of the
386 ecological impact of pest species systems from genomic perspectives will further transform our understanding
387 of sustainable agricultural production. Appropriate experimental designs including careful development of

sampling procedures will be needed for meaningful interpretation that takes advantage of the large volume of genomic data generated by the HTS platform. Our study demonstrates the successful use of multi-primer HTS metabarcoding methods in classical field-ecological studies and for novel cryptic species discovery. The adoption of this approach more widely has the potential to help integrate global biodiversity with genomic (Arribas, Andujar, Bidartondo, & al., 2021) and transform global biosecurity preparedness to impact understanding of ecological network, conservation biology, human health, and food security.

Acknowledgements

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

Data Availability Statement

The data that support the findings of this study are available from GenBank (GenBank accession numbers cited throughout). Raw HTS sequence data can be downloaded from GenBank (SRA xxxx01, xxxx02, xxxx03).

References

Ally, H. M., El Hamss, H., Simiand, C., Maruthi, M. N., Colvin, J., Omongo, C. A., & Delatte, H. (2019). What has changed in the outbreaking populations of the severe crop pest whitefly species in cassava in two decades? *Scientific Reports, 9*. doi:ARTN 14796. 10.1038/s41598-019-50259-0

Arribas, P., Andujar, C., Bidartondo, M. I., & al., (2021). Connecting high-throughput biodiversity inventories: Opportunities for a site-based genomic framework for global integration and synthesis. *Mol Ecol*. doi:10.1111/mec.15797

Bansch, S., Tscharntke, T., Wunschiers, R., Netter, L., Brenig, B., Gabriel, D., & Westphal, C. (2020). Using ITS2 metabarcoding and microscopy to analyse shifts in pollen diets of honey bees and bumble bees along a mass-flowering crop gradient. *Molecular Ecology, 29*(24), 5003-5018. doi:10.1111/mec.15675

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

Brandon-Mong, G. J., Gan, H. M., Sing, K. W., Lee, P. S., Lim, P. E., & Wilson, J. J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bull Entomol Res, 105*(6), 717-727. doi:10.1017/S0007485315000681

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. 10.1186/s12915-016-0321-y

Chi, Y., Pan, L. L., Bouvaine, S., Fan, Y. Y., Liu, Y. Q., Liu, S. S., . . . Wang, X. W. (2020). Differential transmission of Sri Lankan cassava mosaic virus by three cryptic species of the whitefly *Bemisia tabaci* complex. *Virology, 540*, 141-149. doi:10.1016/j.virol.2019.11.013

Crozier, R. H., Crozier, Y. C., & Mackinlay, A. G. (1989). The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Mol Biol Evol, 6*(4), 399-411. doi:10.1093/oxfordjournals.molbev.a040553

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

DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution, 7*. doi:ARTN 302. 10.3389/fevo.2019.00302

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. (2021). On species delimitation, hybridization and population structure of cassava whitefly in Africa. *Scientific Reports 11, 7923; doi: https://doi.org/10.1038/s41598-021-87107-z*

Evans, D. M., Kitson, J. J. N., Lunt, D. H., Straw, N. A., & Pocock, M. J. O. (2016). Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. *Functional Ecology*, *30*(12), 1904-1916. doi:10.1111/1365-2435.12659

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol, 3*(5), 294-299.

Frohlich, D. R., Torres-Jerez, I., Bedford, I. D., Markham, P. G., & Brown, J. K. (1999). A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Molecular Ecology*, *8*(10), 1683-1691. doi:DOI 10.1046/j.1365-294x.1999.00754.x

Gebiola, M., Monti, M. M., Johnson, R. C., Woolley, J. B., Hunter, M. S., Giorgini, M., & Pedata, P. A. (2017). A revision of the *Encarsia pergandiella* species complex (Hymenoptera: Aphelinidae) shows cryptic diversity in parasitoids of whitefly pests. *Systematic Entomology*, *42*(1), 31-59. doi:10.1111/syen.12187

Ghosh, K., & Weiss, L. M. (2009). Molecular diagnostic tests for microsporidia. *Interdiscip Perspect Infect Dis*, *2009*, 926521. doi:10.1155/2009/926521

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

Hajibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLoS One, 14*(9), e0220953. doi:10.1371/journal.pone.0220953

He, X., Gilmore, S. R., Sutherland, T. F., Hajibabaei, M., Miller, K. M., Westfall, K. M., . . . Abbott, C. L. (2021). Biotic signals associated with benthic impact of salmon farms from eDNA metabarcoding of sediments. *Mol Ecol* doi:DOI: 10.1111/mec.15814

Hebert, P.D.N., Braukmann, T.W.A., Prosser, S.W.J. et al. (2018) A Sequel to Sanger: amplicon sequencing that scales. BMC Genomics 19, 219. https://doi.org/10.1186/s12864-018-4611-3

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci U S A, 101*(41), 14812-14817. doi:10.1073/pnas.0406166101

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

Hernández-Suárez, E., Carnero, A., Aguiar, A., Prinsloo, G., LaSalle, J., & Polaszek, A. (2003). Parasitoids of whiteflies (Hymenoptera: Aphelinidae, Eulophidae, Platygastridae; Hemiptera: Aleyrodidae) from the

Macaronesian archipelagos of the Canary Islands, Madeira and the Azores. *Systematics and Biodiversity, 1*(1), 55-108. doi:10.1017/S1477200002001007

Hopkinson, J., Pumpa, S., van Brunschot, S., Fang, G., Frese, M., Tay, W. T., & Walsh, T. (2020). Insecticide resistance status of *Bemisia tabaci* MEAM1 (Hemiptera: Aleyrodidae) in Australia cotton production valleys. *Austral Entomology, 59(1), 202-214*. https://doi.org/10.1111/aen.12436

Kalyebi, A., Macfadyen, S., Hulthen, A., Ocitti, P., Jacomb, F., Tay, W. T., . . . De Barro, P. (2021). Within-season changes in land-use impact pest abundance in smallholder African cassava production systems. *Insects, Revised and rubmitted (9-March, 2021)*.

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

Karut, K., Kaydan, M. B., Tok, B., Doker, I., & Kazak, C. (2015). A new record for *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) species complex of Turkey. *Journal of Applied Entomology, 139*(1-2), 158-160. doi:10.1111/jen.12169

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Klee, J., Tay, W. T., & Paxton, R. J. (2006). Specific and sensitive detection of *Nosema bombi* (Microsporidia : Nosematidae) in bumble bees (*Bombus* spp.; Hymenoptera : Apidae) by PCR of partial rRNA gene sequences. *Journal of Invertebrate Pathology, 91*(2), 98-104. doi:10.1016/j.jip.2005.10.012

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. *BioRxiv* 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

Legg, J. P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R., Bigirimana, S., . . . Brown, J. K. (2014). 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. doi:10.1016/j.virusres.2013.11.018

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., . . . Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool, 10*, 34. doi:10.1186/1742-9994-10-34

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. (2021). Landscape factors and how they influence whitefly pests in cassava fields across East Africa. *Landscape Ecology, 36*(1), 45-67. doi:https://doi.org/10.1007/s10980-020-01099-1

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

Minato, N., Sok, S., Chen, S., Delaquis, E., Phirun, I., Le, V. X., . . . de Haan, S. (2019). Surveillance for Sri Lankan cassava mosaic virus (SLCMV) in Cambodia and Vietnam one year after its initial detection in a single plantation in 2015. *PLoS One, 14*(2), e0212780. doi:10.1371/journal.pone.0212780

Minh, B. Q., Nguyen, M. A., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol, 30*(5), 1188-1195. doi:10.1093/molbev/mst024

Miya, M., Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fisheries Science, 86*(6), 939-970. doi:10.1007/s12562-020-01461-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

O'Mahony, E. M., Tay, W. T., & Paxton, R. J. (2007). Multiple rRNA variants in a single spore of the microsporidian *Nosema bombi*. *Journal of Eukaryotic Microbiology, 54*(1), 103-109. doi:10.1111/j.1550-7408.2006.00232.x

Patil, B. L., & Fauquet, C. M. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Mol Plant Pathol, 10*(5), 685-701. doi:10.1111/j.1364-3703.2009.00559.x

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., . . . de Vargas, C. (2012). CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *Plos Biology, 10*(11). doi:ARTN e1001419. 10.1371/journal.pbio.1001419

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

Popovic, A., Bourdon, C., Wang, P. W., Guttman, D. S., Voskuijl, W., Grigg, M. E., . . . Parkinson, J. (2018). Design and application of a novel two-amplicon approach for defining eukaryotic microbiota. *Microbiome, 6*. doi:ARTN 228. 10.1186/s40168-018-0612-3

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

Roehrdanz, R. L. (1993). An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. *Insect Mol Biol, 2*(2), 89-91. doi:10.1111/j.1365-2583.1993.tb00129.x

607 Roopa, H. K., Asokan, R., Rebijith, K. B., Hande, R. H., Mahmood, R., & Kumar, N. K. K. (2015). Prevalence of a
608 new genetic group, MEAM-K, of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Karnataka, India, as
609 evident from mtCOI sequences. *Florida Entomologist, 98*(4), 1062-1071.

610 Rossitto De Marchi, B., Kinene, T., Mbora Wainaina, J., Krause-Sakate, R., & Boykin, L. (2018). Comparative
611 transcriptome analysis reveals genetic diversity in the endosymbiont *Hamiltonella* between native and
612 exotic populations of *Bemisia tabaci* from Brazil. *PLoS One, 13*(7), e0201411.
613 doi:10.1371/journal.pone.0201411

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

617 Shatters, R. G., Jr., Powell, C. A., Boykin, L. M., Liansheng, H., & McKenzie, C. L. (2009). Improved DNA
618 barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci*
619 biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *J Econ Entomol,*
620 *102*(2), 750-758. doi:10.1603/029.102.0236

621 Shelomi, M., & Chen, M. J. (2020). Culturing-Enriched Metabarcoding Analysis of the *Oryctes rhinoceros* Gut
622 Microbiome. *Insects, 11*(11). doi:ARTN 782. 10.3390/insects11110782

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

626 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, Weighting, and
627 Phylogenetic Utility of Mitochondrial Gene-Sequences and a Compilation of Conserved Polymerase Chain-
628 Reaction Primers. *Annals of the Entomological Society of America, 87*(6), 651-701. doi:DOI
629 10.1093/aesa/87.6.651

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

634 Sow, A., Brevault, T., Benoit, L., Chapuis, M. P., Galan, M., d'acier, A. C., . . . Haran, J. (2019). Deciphering host-
635 parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports, 9*.
636 doi:ARTN 3646. 10.1038/s41598-019-40243-z

637 Srivathsan, A., Lee, L., Katoh, K. et al. ONTbarcoder and MinION barcodes aid biodiversity discovery and
638 identification by everyone, for everyone. BMC Biol 19, 217 (2021). https://doi.org/10.1186/s12915-021-
639 01141-x

640 Sseruwagi, P., Sserubombwe, W. S., Legg, J. P., Ndunguru, J., & Thresh, J. M. (2004). Methods of surveying the
641 incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a
642 review. *Virus Res, 100*(1), 129-142. doi:10.1016/j.virusres.2003.12.021

643 Stefanni, S., Stankovic, D., Borme, D., de Olazabal, A., Juretic, T., Pallavicini, A., & Tirelli, V. (2018). Multi-marker
644 metabarcoding approach to study mesozooplankton at basin scale. *Scientific Reports, 8*. doi:ARTN 12085.
645 10.1038/s41598-018-30157-7

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

649 Tay, W. T., Elfekih, S., Court, L., Gordon, K. H., & De Barro, P. J. (2016). Complete mitochondrial DNA genome of
650 *Bemisia tabaci* cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). *Mitochondrial DNA Part A,*
651 *27*(2), 972-973. doi:10.3109/19401736.2014.926511

652 Tay, W. T., Elfekih, S., Court, L. N., Gordon, K. H. J., Delatte, H., & De Barro, P. J. (2017). The Trouble with
653 MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive *Bemisia tabaci*
654 (Hemiptera: Aleyrodidae) Cryptic Species Complex. *Genome Biology and Evolution, 9*(10), 2732-2738.
655 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. 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 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., O'Mahony, E. M., & Paxton, R. J. (2005). Complete rRNA gene sequences reveal that the microsporidium *Nosema bombi* infects diverse bumblebee (Bombus spp.) hosts and contains multiple polymorphic sites. *Journal of Eukaryotic Microbiology, 52*(6), 505-513. doi:10.1111/j.1550-7408.2005.00057.x

Thao, M. L., & Baumann, P. (2004). Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl Environ Microbiol, 70*(6), 3401-3406. doi:10.1128/AEM.70.6.3401-3406.2004

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

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res, 44*(W1), W232-235. doi:10.1093/nar/gkw256

Ueda, S., Kitamura, T., Kijima, K., Honda, K. I., & Kanmiya, K. (2009). Distribution and molecular characterization of distinct Asian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Japan. *Journal of Applied Entomology, 133*(5), 355-366. doi:10.1111/j.1439-0418.2008.01379.x

Velasquez, J. N., Carnevale, S., Guarnera, E. A., Labbe, J. H., Chertcoff, A., Cabrera, M. G., & Rodriguez, M. I. (1996). Detection of the microsporidian parasite *Enterocytozoon bieneusi* in specimens from patients with AIDS by PCR. *Journal of Clinical Microbiology, 34*(12), 3230-3232.

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

Wang, H. L., Xiao, N., Yang, J., Wang, X. W., Colvin, J., & Liu, S. S. (2016). The complete mitochondrial genome of *Bemisia afer* (Hemiptera: Aleyrodidae). *Mitochondrial DNA A DNA Mapp Seq Anal, 27*(1), 98-99. doi:10.3109/19401736.2013.873921

Wang WY, Srivathsan A, Foo M, Yamane SK, Meier R. (2018) Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. Molecular Ecology Resources 18(3), 490-501.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci, 360*(1462), 1847-1857. doi:10.1098/rstb.2005.1716

Weitemier, K., Penaluna, B. E., Hauck, L. L., Longway, L. J., Garcia, T., & Cronn, R. (2021). Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding. . *Mol Ecol*. doi:10.1111/mec.15811

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a Natural System of Organisms - Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America, 87*(12), 4576-4579. doi:DOI 10.1073/pnas.87.12.4576

Woodcock, T. S., Boyle, E. E., Roughley, R. E., Kevan, P. G., Labbee, R. N., Smith, A. B., . . . Adamowicz, S. J. (2013). The diversity and biogeography of the Coleoptera of Churchill: insights from DNA barcoding. *BMC Ecol, 13*, 40. doi:10.1186/1472-6785-13-40

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

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

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

711 Xu, J. P. (2016). Fungal DNA barcoding. *Genome, 59*(11), 913-932. doi:10.1139/gen-2016-0046

712