

The Trouble with MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive *Bemisia tabaci* (Hemiptera: Aleyrodidae) Cryptic Species Complex

Wee Tek Tay^{1,*}, Samia Elfekih¹, Leon N. Court¹, Karl H.J. Gordon¹, H  l  ne Delatte², and Paul J. De Barro³

¹CSIRO, Black Mountain Science and Innovation Park, Acton, Australia

²CIRAD, UMR PVBMT, La R  union, France

³CSIRO, Ecosciences Precinct, Brisbane, Queensland, Australia

*Corresponding author: E-mail: weetek.tay@csiro.au.

Accepted: August 29, 2017

Data deposition: This project has been deposited at NCBI GenBank under the accession numbers KY951447, KY951448, KY951449, KY951450, KY951451, KY951452, KY951453, KY951454.

Abstract

Molecular species identification using suboptimal PCR primers can over-estimate species diversity due to coamplification of nuclear mitochondrial (NUMT) DNA/pseudogenes. For the agriculturally important whitefly *Bemisia tabaci* cryptic pest species complex, species identification depends primarily on characterization of the mitochondrial DNA cytochrome oxidase I (mtDNA COI) gene. The lack of robust PCR primers for the mtDNA COI gene can undermine correct species identification which in turn compromises management strategies. This problem is identified in the *B. tabaci* Africa/Middle East/Asia Minor clade which comprises the globally invasive Mediterranean (MED) and Middle East Asia Minor I (MEAM1) species, Middle East Asia Minor 2 (MEAM2), and the Indian Ocean (IO) species. Initially identified from the Indian Ocean island of R  union, MEAM2 has since been reported from Japan, Peru, Turkey and Iraq. We identified MEAM2 individuals from a Peruvian population via Sanger sequencing of the mtDNA COI gene. In attempting to characterize the MEAM2 mitogenome, we instead characterized mitogenomes of MEAM1. We also report on the mitogenomes of MED, AUS, and IO thereby increasing genomic resources for members of this complex. Gene synteny (i.e., same gene composition and orientation) was observed with published *B. tabaci* cryptic species mitogenomes. Pseudogene fragments matching MEAM2 partial mtDNA COI gene exhibited low frequency single nucleotide polymorphisms that matched low copy number DNA fragments (<3%) of MEAM1 genomes, whereas presence of internal stop codons, loss of expected stop codons and poor primer annealing sites, all suggested MEAM2 as a pseudogene artifact and so not a real species.

Key words: invasive pest, mitogenome, pseudogene, NUMT, high throughput -sequencing.

Introduction

The use of single-gene based DNA barcoding to resolve species boundaries for cryptic species presents a special challenge. The resolution of such morphologically identical species based on a single gene sequence alignment is only possible if that gene sequence is unambiguously correct and corresponds to what is expected in every case analyzed. For many such taxonomic exercises, mitochondrial genes and primarily the mitochondrial cytochrome oxidase I gene (mtDNA COI) barcode sequence have been selected (e.g., Alam et al. 2015; Leys et al. 2016). The same applies to other

mitochondrial sequences such as the cytochrome oxidase II gene (mtDNA COII) (e.g., Sunnucks et al. 2000) and cytochrome *b* gene (cyt *b*) (e.g., Mundy et al. 2000), as well as nuclear DNA markers (i.e., nuclear 18S and 28S rRNA gene regions, e.g., Jorger and Schrodler 2013; microsatellite DNA markers, e.g., Cheng et al. 2013).

One significant challenge faced in the delimitation of otherwise indistinguishable species using mtDNA COI data sets is the possible presence of nuclear mitochondrial DNA pseudogenes (NUMTs) (Bensasson et al. 2001). PCR products derived from NUMTs are often a result of poor PCR primer efficiencies

(Moulton et al. 2010; Lobo et al. 2013; Tay et al. 2017). When they are treated as authentic mtDNA genes, failure to identify them is likely to lead to inaccurate phylogenetic inferences due to differences in divergence times between NUMTs and genuine mtDNA genes (Sunnucks et al. 2000; Bensasson et al. 2001). Numerous methodologies are available to assist with the identification of NUMTs (reviewed in Bensasson et al. 2001) and include identifying non-functionality of the gene fragment through the presence of stop codons within protein coding gene regions. Despite this, NUMTs can be overlooked (e.g., Boykin et al. 2007; Karut et al. 2015) when stop codons are found outside the target gene region and this is a particular challenge when the target sequences are short.

Arthropod pests of economic importance, such as those infesting stored grain products (Tay et al. 2016a), vectors of plant or animal pathogens, for example, cassava brown streak virus (Maruthi et al. 2005), tomato yellow leaf curl virus (Ghanim et al. 1998), blue tongue virus (Tabachnick 1996; De Liberato et al. 2005), and parasites, for example, varroa mites (Anderson and Trueman 2000), require accurate species identification in order to better understand population structure of these pests, the interpretation of disease outbreaks, detection of vector-host association (e.g., Tabachnick 1996), and incursion patterns (Tay et al. 2016b). Further, biosecurity preparedness (i.e., early detection) and incursion response strategies as part of national border protection, require unambiguous knowledge of species status (e.g., Armstrong and Ball 2005; Collins et al. 2012; Tay et al. 2016b).

In the whitefly *Bemisia tabaci* pest species complex, numerous species belonging to at least 11 (De Barro et al. 2011) sister clades have been proposed based on the partial mtDNA COI gene, the 657 bp 3' end of the gene (Boykin et al. 2013; Lee et al. 2013). Of these clades, the African/Middle East/Asia Minor clade is of special interest, as it contains the two most invasive members of the complex, Mediterranean (MED, the "true" *B. tabaci* [Tay et al. 2012]) and Middle East Asia Minor 1 (MEAM1). This clade also contains two other species, Middle East Asia Minor 2 (MEAM2) and Indian Ocean (IO). Whilst IO is not known to be invasive, MED and MEAM1 are globally wide-spread and vectors of highly damaging plant viruses (reviewed in De Barro et al. 2011), whereas MEAM2 has increasingly been detected across globally disparate locations such as Japan (Ueda et al. 2009; AB308110), Peru (this study), Iraq (KX679576; collected in 2015), Turkey (Karut et al. 2015, sequences KK103B, KK104A, KK104B), and Egypt (FJ939600, FJ939602), since its initial detection in the Indian Ocean island of Réunion (Delatte et al. 2005; AJ550177). The incursion pathways of MED and MEAM1 are linked to the worldwide global trade in ornamental plants (Cheek and Macdonald 1994; Dalton 2006), however factors underlying the spread of MEAM2 are less certain although detection frequencies have increased in recent times since its initial report (Delatte et al. 2005).

Molecular characterizations using the whole mitogenome have been carried out for only one of the four invasive clade species—that of MED, although other *Bemisia* species from the complex and *Bemisia* "JpL" species have also been reported (Baumann 2004; Wang et al. 2013; Tay et al. 2016, 2017). In this study, we characterized the complete mitogenome of MEAM2 using high-throughput sequencing methods, and in the process ascertained the molecular genetic basis for the species delimitation of MEAM2. This effort also enabled the molecular characterization of two remaining "invasive clade" *B. tabaci* cryptic species (i.e., MEAM1, IO) draft mitogenomes, as well as the draft mitogenome of the Australia *B. tabaci* (previously biotype "AN", De Barro et al. 2011) to be characterized via the high-throughput sequencing method. We assessed and discussed the impact of NUMT on phylogenetic inferences on the cryptic *B. tabaci* species complex.

Materials and Methods

Bemisia tabaci Samples, gDNA Extraction, PCR, and NGS

Five individuals of *Bemisia* whiteflies from a single Peruvian population, collected on August 14, 2000 from Cañete Valley (GenBank KY951453, KY951454, KX234912, KX234913, KX234914), four from Ouagadougou, Burkina Faso (KX234908, KX234909, KX234910, KX234911), two Australian *Bemisia* whiteflies (mtDNA COI matched (100%) MEAM1 (DQ174535; Hsieh et al. 2006) from Bundaberg, Australia), and five from Réunion (KX234868, KX234869, KX234870, KX234871, KX234872) were analyzed via standard PCR and Sanger sequencing procedures (e.g., see Dinsdale et al. 2010). Sanger sequencing was carried out at the John Curtin School of Medical Research Biological Resource Facility at the National University of Australia, Canberra. Sanger sequence trace files were assembled using Staden Pregap4 and Gap4 programs (Staden et al. 2000), and species status determined using BlastN searches against the publicly available *B. tabaci* mtDNA COI database <<http://dx.doi.org/10.4225/08/50EB54B6F1042>> (last accessed September 6, 2017). All genomic DNA (gDNA) extractions were performed using the Qiagen DNeasy Blood and Tissue kit (Cat. # 69506), including the optional RNase A treatment (Qiagen, Cat. # 19101). Individually extracted and purified gDNA samples were eluted in 25 µl of Qiagen buffer EB (Cat. # 19086) and quantified using a Qubit 2.0 Fluorometer and the Qubit dsDNA High Sensitivity DNA Assay kit (ThermoFisher Scientific, Cat # Q32854).

The gDNA from three of the five Peruvian whitefly specimens (KX234913, KX234914, KY951454) were each made into separate NGS gDNA libraries using the protocol of Tay et al. (2016) and sequenced using the Illumina MiSeq sequencer. To better understand the potential genomic origins

of MEAM2 COI haplotypes and hence its species status, we further prepared separate Illumina MiSeq libraries of a single individual from each of the three species (i.e., MED, IO, MEAM1) known to be also present in Réunion Island (Delatte et al. 2005). These included one Réunion individual from an “IO” population, one Burkina Faso individual from a MED population, and one MEAM1 individual from an Australian population. The high throughput sequencing gDNA library preparation method followed the Illumina Nextera XT DNA library preparation guide (Part # 15031942 Rev. D, September 2014).

Briefly, 1.5 ng samples of gDNA were tagged (i.e., tagged and fragmented by the Nextera XT transposome), followed by limited PCR cycles (to add unique dual index barcodes for sample tracking and Illumina adapters for cluster formation). The amplified libraries were sized selected and purified using the Beckman Coulter AMPure XP system (Bead to DNA ratio of 0.7) and eluted in 28 μ l of Qiagen buffer EB (Cat. # 19086). The purified libraries were then quantified by Qubit dsDNA High Sensitivity DNA Assay as above, their average fragment size estimated using the Agilent 2200 TapeStation and High Sensitivity D1000 screentape (Cat # 5067-5585) and then normalized to a final concentration of 4 nM. The Nextera XT gDNA libraries were pooled, diluted to a final concentration of 11 pM (with 5% spike-in of Illumina Phi X Control v3 library [Cat # FC-110-3001]) and sequenced on the Illumina MiSeq sequencer. The draft mitogenomes were individually assembled using the Asia I mitogenome (GenBank KJ778614) of Tay et al. (2016) as the reference genome within the genomic analysis software Geneious 8.1.9 (Biomatters Ltd., NZ). To confirm the circular nature of the mitogenomes we individually assembled the intergenic region between the NAD2 and COI genes, starting with either the NAD2 or the COI gene and allowing the assembly to bridge across to the adjacent gene.

Mitogenome Annotation and Identification of NUMTs

Assembled mitogenomes were annotated using MITOS (Bernt 2013) prior to manual readjustment within Geneious 8.1.9 to identify potential stop codons in all coding sequences (KY951447, KY951448, KY951449, KY951450, KY951451, KY951452). Assembled draft mitogenomes were reconfirmed for species identity by Blastn searches of the partial (657 bp) mtDNA COI gene region against the GenBank DNA database. To assess the impact of NUMTs in misidentification of MEAM1 as MEAM2, a Peruvian MEAM2 mtDNA COI sequence detected (KX234914), as well as published sequences (Delatte et al. 2005; Ueda et al. 2009; Karut et al. 2015); were used as template reference sequences and assessed for frequencies of SNPs detected at the respective genomic regions/nucleotide positions in the three species (i.e., MEAM1, MED, and IO) known to be present in countries that have also reported MEAM2 (e.g., Reunion, Turkey,

Japan, Peru, and Iraq). We also visually identify MiSeq generated DNA fragments that uniquely matched SNP patterns of MEAM2 partial mtDNA COI regions to determine the effects on the amino acid translational processes.

Results and Discussion

Our results supported the notion that MEAM2 partial mtDNA COI sequences reported to-date are likely to be NUMTs. We also generated and characterized mitogenomes of four (MED, MEAM1, IO and AUS) *B. tabaci* cryptic species from single individuals, of which the complete mitogenomes of three species (MEAM1, IO and AUS) are here reported for the first time. Based on our initial Sanger sequencing, two individuals from the Australian collection were identified as MEAM1. However, the third individual analyzed via NGS from the same collection was identified as belonging to a different member of the complex, AUS (657 bp mtDNA COI partial gene matched 100% sequence identity to Bundaberg, Australia [GU086328]), indicating that the collection consisted of both MEAM1 and AUS.

For the randomly selected Réunion individual as well as the Burkina Faso individual, we obtained the expected mitogenomes of IO (657 bp mtDNA COI partial gene matched 100% sequence identity of a Madagascan IO [AJ550171]) and MED (partial mtDNA COI gene (657 bp) shared 100% sequence identity to MED from Sudan [DQ133378]), respectively. From the three Peruvian individuals that were expected to be MEAM2 (i.e., KY951454; KX234913, and KX234914) on the basis of the Sanger sequence derived mtDNA COI partial gene, we instead obtained MEAM1 mitogenomes, as confirmed via partial mtDNA COI gene comparison with published sequences (KY951452 and KX234913 [nt782-1, 439] 100% sequence identity to MEAM1 from Arizona, USA [HM070411]; and KX234914 [nt782-1, 439] = 99% sequence identity to MEAM1 from, e.g., Florida, USA [GU086340]). MEAM1 had previously been argued to represent a separate *Bemisia* species from *B. tabaci* based on behavioral, morphological, and genetic differences (e.g., Bellows et al. 1994; Perring et al. 1992, 1993) and was subsequently named *B. argentifolii* (Bellows et al. 1994). Thao et al. (2004) provided partial regions (i.e., Cyt *b*-COIII, 4, 796 bp; GenBank AY521257) of the *B. argentifolii* mitogenome, however the complete mitogenome of MEAM1/*B. argentifolii* had not been published. Pairwise sequence comparisons between AY521257 and our reported MEAM1 mitogenomes identified high levels of sequence similarity (99.82% identity) with the corresponding *B. argentifolii* mitogenome region, whereas similarity between MED, IO, and AUS mitogenome regions were much lower, at 92.52%, 91.51%, and 80.16% sequence identity, respectively (data not shown).

Sequencing of these gDNA libraries generated between 2.15 and 28.96 million paired-end (PE) sequences (table 1), from which 10,738 to 131,328 PE sequences were assembled

Table 1

Summary Statistics of MiSeq Sequence Data from *Bemisia tabaci* Cryptic Species of Indian Ocean (IO, KY951448), Mediterranean (MED, KY951447), Middle East-Asia Minor 1 (MEAM1, KY951449, KY951450, KY951452), and Australia (AUS, KY951451)^a

Species	Total PE-seq	MTG PE-seq	Average COI Coverage (\pm s.d.)	Mitogenome Lengths ^b	GenBank
MEAM1	6,514,260	13,986	166.9 \pm 16.4 s.d.	15,666	KY951450
MEAM1	28,964,958	131,328	990.3 \pm 163.4 s.d.	15,531	KY951449
MEAM1	6,663,490	12,176	84.0 \pm 21.0 s.d.	15,526	KY951452
MED	2,157,716	11,380	126.7 \pm 28.5 s.d.	15,631	KY951447
IO	3,842,616	10,738	118.3 \pm 22.1 s.d.	15,626	KY951448
AUS	4,981,182	15,438	173.9 \pm 25.3 s.d.	15,686	KY951451
MED	N/A	N/A	N/A	15,632	JQ906700
ASIA I	N/A	N/A	N/A	15,210	KJ778614
ASIA II-7	N/A	N/A	N/A	15,515	KX714967

^aThe overall published draft mtDNA genomes of *B. tabaci* cryptic species ranged between 15,210 in *B. tabaci* Asia I (KJ778614) to 15,686 in *B. tabaci* AUS (KY951451).

^bMitogenome lengths from this study are putative due to the difficulty of assembling complete mitogenomes based on short read DNA sequences as obtained from the Illumina MiSeq sequencing method. N/A (not applicable)—these are either from published data or not available. Average COI coverage information included average sequence reads across the whole mtDNA COI gene and standard deviation (s.d.), as calculated using Geneious version 8.1.9.

to generate complete mitogenomes in IO, MEAM1, MED, and AUS (table 1). We identified low copy genome fragments through the Illumina MiSeq sequencing platform in MEAM1 individuals that matched unique MEAM2 SNPs (fig. 1A and B). Fragments of gDNA representing the MEAM2 partial mtDNA COI haplotypes also identified the presence of premature stop codons within these low copy number DNA fragments in regions of the mtDNA COI gene, as well as the loss of the expected stop codon at the C-terminal region of the mtDNA COI gene (fig. 1A and 1B). Corresponding SNP frequencies across DNA fragments generated from high-throughput sequencing, and that potentially represented NUMTs within the 657 bp mtDNA COI partial gene region, were detected at very low frequencies (supplementary table 1a, Supplementary Material online), again supporting the notion that NUMTs which had resulted in the misidentification of “MEAM2” sequences, were present as low copy DNA fragments. At the corresponding nucleotide positions between a randomly selected MEAM1 sequence from GenBank and compared against the MEAM1 DNA fragments generated from the high-throughput sequencing library, SNPs detected at nucleotide positions that corresponded to those in MEAM2 were generally observed at highest frequencies (supplementary tables 1b, 2b, 3b, 4b, 5b, and 6b, Supplementary Material online). For MEAM2 when compared with MED and IO, there were no particular SNP frequency patterns (supplementary tables 1c and d; 2c and d; 3c and d; 4c and d; 5c and d; and 6c and d, Supplementary Material online). Contrasting this, SNPs within suspected MEAM2 sequences (i.e., Japan AB308110, the Peruvian MEAM2 sequence (KX234913), four Turkish MEAM2 haplotypes (Karut et al. 2015) were consistently of the lower frequencies (supplementary tables 1a, 2a, 3a, 4a, 5a, and 6a, Supplementary Material online). Characterization of the MEAM1 mitogenomes therefore supported the hypothesis that the MEAM2 sequences were likely associated with low copy DNA fragments from the MEAM1

genome and were most likely either PCR artifacts such as DNA polymerase-introduced errors or nuclear mitochondrial DNA (e.g., NUMTs).

A further piece of supporting evidence that MEAM2 belonged to NUMT was from the recently assembled MEAM1 draft genome (Chen et al. 2016), in which an unknown protein coding gene predicted to be cytochrome c oxidase subunit 1-like mRNA (XM_019045089.1) was identified; it shared 99% sequence homologies with the Peruvian KX234914 MEAM2 partial COI gene. Within this COI-like-mRNA sequence four internal stop codons were identified and subsequently corrected (i.e., modifications involving substitutions of four bases at four genomic stop codons were introduced to the sequence of the model RefSeq protein relative to its source genomic sequence so as to represent the inferred coding sequences [GenBank Locus XM_019045089, 1,632 bp mRNA linear INV November 9, 2016; accessed January 5, 2017]). NUMTs are widespread in all eukaryotic organisms, can both be difficult to detect and introduce bias in the estimation of species diversity and DNA barcoding analyses (reviewed in Hazkani-Covo et al. 2010). Our analysis therefore supported the presence of only three species (i.e., MED, MEAM1, and IO) within the current invasive *B. tabaci* clade (Asia/Middle East/Asia Minor), and indicated that MEAM2 was a NUMT artifact. With increasing molecular characterization of global *B. tabaci* cryptic species complex, new species may be identified which could alter the current *B. tabaci* cryptic species phylogeny and also ultimately the number of species within the invasive *B. tabaci* clade.

Our efforts to understand species composition and to ascertain the spread of invasive *B. tabaci* based on limited individuals have initially identified MEAM1 in the Australian samples from Bundaberg, Queensland. When additional individuals were sampled in high-throughput sequencing we instead obtained the native AUS species. From the Peruvian

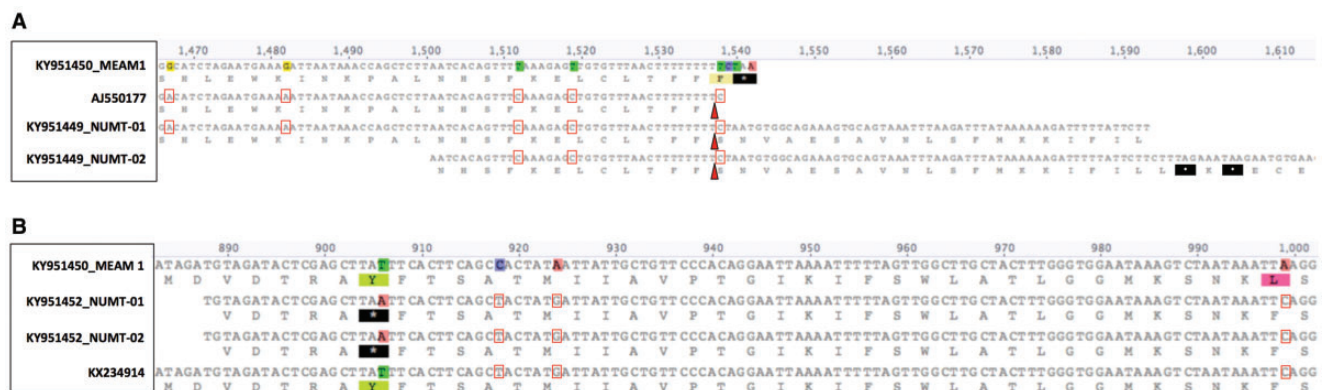


Fig. 1.—Examples of sequence alignments between *Bemisia tabaci* “Peru” MEAM1 mtDNA COI haplotype gene region, published “MEAM2” mtDNA COI gene region, and NGS candidate NUMT sequences identified from KX234913, KY951449, and the KY951452 individuals. (A) C-terminal region of a Peruvian MEAM1 *B. tabaci* (KY951450) mtDNA COI gene region showing putative stop codon (black shaded “*” symbol), as well as the *B. tabaci* “MEAM2” haplotype from Japan (AJ550177), and examples NGS candidate NUMT sequences from the Peruvian *B. tabaci* MEAM1 (KX234914) with matching SNPs (indicated by red boxes) that matched the Japan MEAM2 haplotype (AJ550177). Deletion of a “T” base (indicated by red triangle) resulted in a frameshift mutation and the loss of the putative mtDNA COI gene stop codon. (B) Internal stop codons (at positions 904S906) detected in candidate NUMT sequences (KY951452_NUMT-01, KY951452_NUMT-02) from the Peruvian MEAM1 individual (KY951452) MiSeq generated DNA fragments when compared with the Peruvian *B. tabaci* MEAM1 (KX234914) mtDNA COI gene. Stop codons detected in NUMT sequences were the result of a single nucleotide base change at position 906 from a “T” to an “A.” Candidate NUMT sequences were also compared with the Peruvian MEAM2 haplotype (KY951454) obtained via PCR and Sanger sequencing of the same MEAM1 individual (KY951452). Nucleotide positions based on the mtDNA COI gene are provided. Amino acid translation based on the invertebrate mitochondrial genetic codes (Translational Table_5). Significant changes between amino acids are highlighted.

individuals initially identified as MEAM2 based on partial mtDNA COI gene using suboptimal primers, high-throughput sequencing have also resulted in MEAM1 mitogenomes being assembled instead. This exercise highlighted the importance of analyzing an adequate number of individuals from a collection and the impact suboptimal PCR primers can have on estimating species composition. These included misidentification of species composition complexity at the population level, and minimizing valuable resources being misdirected to monitor for incursion of nonexistent species, both of which can have profound impacts in terms of border biosecurity responses (e.g., either missing or misidentifying species of biosecurity concern).

Several published studies (e.g., Delatte et al. 2005; Ueda et al. 2009; Karut et al. 2015) have used various non *Bemisia* “universal” PCR primers such as C1-J-2195/L2-N-3014; C1-J-2195/R-BQ-2819; C1-J-2195/tRNA-1576 (Simon et al. 1994; Frohlich et al. 1999; Tsagkarakou et al. 2007; Chu et al. 2011) and we suspect, factors such as reduced annealing site specificity (supplementary table 7, Supplementary Material online) are contributing to the coamplification of NUMTs. Previous studies reporting the detection of MEAM2, using the C1-J-2195 forward non *Bemisia* “universal” primer was a common factor. This primer, originally named “COI-RLR”, was developed by Roehrdanz (1993) from the *Apis mellifera* COI gene (Crozier and Crozier 1993), and was shown to amplify some Lepidoptera, Coleoptera, Diptera, and Hymenoptera, but with unknown efficacies for Hemiptera (Simon et al. 1994) to which *Bemisia* belongs.

Various *Bemisia* species’ complete mitogenomes are now available (MEAM1, IO, AUS (this study), MED (Wang et al. 2013), Asia I (Tay et al. 2016), Asiall_7 (originally identified as *B. emiliae*, but synonymised with *B. tabaci* in 1957, Tay et al. 2017). Direct comparison of primer-binding site efficacies between the C1-J-2195 24-mer oligonucleotide and the intended COI gene target site in these species identified poor primer efficacies that ranged between 33.3% and 45.8% for MEAM1, MED, IO, AUS, Asia I, and Asiall_7 (supplementary table 7, Supplementary Material online). *B. tabaci* cryptic species mtDNA COI sequences as generated using the C1-J-2195 primer should therefore be treated with extra caution. The sequencing of full mitogenomes in *B. tabaci* whiteflies can be achieved from single adults or nymphs (Tay et al. 2016, 2017; this study) and will significantly contribute to development of *B. tabaci* species-specific primers, although standardization of PCR-primers would be of benefit to the *B. tabaci* research community (Elfekih et al. 2017).

We have shown the consequence of pseudogenes on species delimitation within the *B. tabaci* cryptic complex, through direct and active searching of genomic fragments obtained from high-throughput sequencing against suspected NUMTs of the “MEAM2” haplotypes. Studies investigating the species status within the *B. tabaci* complex have, to-date, relied largely on the C1-J-2195 primer and have generated a large volume of haplotype data across the breadth of the *B. tabaci* complex. These haplotypes, currently >5,100 sequences (GenBank accessed March 17, 2017), will likely contain other unidentified pseudogenes. Future studies focusing on the

phylogenetic relationships within the complex will need to be mindful of NUMTs and will require careful treatment of data so as to avoid over-interpretation of *B. tabaci* phylogeny and species status.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Chris Hardy, Tom Walsh, and William James (CSIRO) provided valuable discussion during the course of this study. S.E. was supported by a CSIRO OCE Post Doctoral Fellowship (R-4546-1). W.T.T. and S.E. acknowledged support from CSIRO Health and Biosecurity (R-8681-1). Peruvian *B. tabaci* was kindly provided by P.K. Anderson to PJDB. The authors declare no conflict of interest.

Authors' Contribution

Project design: W.T.T., S.E., L.N.C., H.D., K.H.J.G., P.J.D.B. Laboratory work: W.T.T., L.C., S.E. Data analysis: W.T.T., S.E., Manuscript preparation: W.T.T., S.E., L.N.C., K.H.J.G., H.D., P.J.D.B. All authors have read and agreed to the manuscript.

Literature Cited

- Alam MM, Westfall KM, Palsson S. 2015. Mitochondrial DNA variation reveals cryptic species in *Fenneropenaeus indicus*. *Bull Mar Sci*. 91(1): 15–31.
- Anderson DL, Trueman JW. 2000. *Varroa jacobsoni* (Acari: Varroidea) is more than one species. *Exp Appl Acarol*. 24(3): 165–189.
- Armstrong KF, Ball SL. 2005. DNA barcodes for biosecurity: invasive species identification. *Philos Trans R Soc Lond B Biol Sci*. 360(1462): 1813–1823.
- Baumann L. 2004. Sequence analysis of DNA fragments from the genome of the primary endosymbiont of the whitefly *Bemisia tabaci*. *Curr Microbiol*. 48(1): 77–81.
- Bellows TS, Perring TM, Gill RJ, Headrick DH. 1994. Description of a species of *Bemisia* (Homoptera, Aleyrodidae). *Ann Entomol Soc Am*. 87(2): 195–206.
- Bensasson D, Zhang D, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol*. 16(6): 314–321.
- Bernt M. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 69(2): 313–319.
- Boykin LM, Bell CD, Evans G, Small I, De Barro PJ. 2013. Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC Evol Biol*. 13: 228.
- Boykin LM, et al. 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol*. 44(3): 1306–1319.
- Cheek S, Macdonald O. 1994. Statutory controls to prevent the establishment of *Bemisia tabaci* in the United-Kingdom. *Pesticide Sci*. 42: 135–137.
- Chen W, et al. 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 Bioinformatics* 14(1): 110.
- Cheng S, Lee CT, Wan MN, Tan SG. 2013. Microsatellite markers uncover cryptic species of *Odontotermes* (Termitoidea: Termitidae) from Peninsular Malaysia. *Gene* 518(2): 412–418.
- Chu D, Gao CS, De Barro P, Wan FH, Zhang YJ. 2011. Investigation of the genetic diversity of an invasive whitefly (*Bemisia tabaci*) in China using both mitochondrial and nuclear DNA markers. *Bull Entomol Res*. 101(4): 467–475.
- Collins RA, et al. 2012. Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. *PLoS One* 7(1): e28381.
- Crozier RH, Crozier YC. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133(1): 97–117.
- Dalton R. 2006. Whitefly infestations: the Christmas invasion. *Nature* 443(7114): 898–900.
- De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. 2011. *Bemisia tabaci*: a statement of species status. *Annu Rev Entomol*. 56: 1–19.
- De Liberato C, et al. 2005. Identification of *Culicoides obsoletus* (Diptera: Ceratopogonidae) as a vector of bluetongue virus in central Italy. *Vet Rec*. 156(10): 301–304.
- Delatte H, et al. 2005. A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the southwest Indian Ocean. *Bull Entomol Res*. 95(1): 29–35.
- Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. 2010. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Ann Entomol Soc Am*. 103(2): 196–208.
- Elfekih S, Tay WT, Gordon K, Court L, De Barro P. 2017. Standardised molecular diagnostic tool for the identification of cryptic species within the *Bemisia tabaci* complex. *Pest Manag Sci*. doi: 10.1002/ps.4676.
- Frohlich DR, Torres-Jerez II, Bedford ID, Markham PG, Brown JK. 1999. A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol Ecol*. 8(10): 1683–1691.
- Ghanim M, Morin S, Zeidan M, Czosnek H. 1998. Evidence for transovarial transmission of tomato yellow leaf curl virus by its vector, the whitefly *Bemisia tabaci*. *Virology* 240(2): 295–303.
- Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet*. 6(2): e1000834.
- Hsieh CH, Wang CH, Ko CC. 2006. Analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex and distribution in Eastern Asia based on mitochondrial DNA markers. *Ann Entomol Soc Am*. 99: 768–775.
- Jorger KM, Schrod M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front Zool*. 10(1): 59.
- Karut K, Kaydan MB, Tok B, Döker I, Kazak C. 2015. A new record for *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) species complex of Turkey. *J Appl Entomol*. 139(1–2): 158–160.
- Lee W, Park J, Lee GS, 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.
- Leys M, Keller I, Räsänen K, Gattoliat J-L, Robinson CT. 2016. Distribution and population genetic variation of cryptic species of the Alpine mayfly *Baetis alpinus* (Ephemeroptera: Baetidae) in the Central Alps. *BMC Evol Biol*. 16: 77.
- Lobo J, et al. 2013. Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC Ecol*. 13(1): 34.
- Maruthi MN, et al. 2005. Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). *J Phytopathol*. 153(5): 307–312.
- Moulton MJ, Song HJ, Whiting MF. 2010. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a

- case study from Orthoptera (Arthropoda: Insecta). *Mol Ecol Resour.* 10(4): 615–627.
- Mundy NI, Pissinatti A, Woodruff DS. 2000. Multiple nuclear insertions of mitochondrial cytochrome *b* sequences in callitrichine primates. *Mol Biol Evol.* 17(7): 1075–1080.
- Perring TM, Cooper A, Kazmer DJ. 1992. Identification of the poinsettia strain of *Bemisia tabaci* (Homoptera, Aleyrodidae) on broccoli by electrophoresis. *J Econ Entomol.* 85(4): 1278–1284.
- Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TS Jr. 1993. Identification of a whitefly species by genomic and behavioral studies. *Science* 259(5091): 74–77.
- Roehrdanz RL. 1993. An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. *Insect Mol Biol.* 2(2): 89–91.
- Simon C, et al. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Ann Entomol Soc Am.* 87(6): 651–701.
- Staden R, Beal KF, Bonfield JK. 2000. The Staden package, 1998. *Methods Mol Biol.* 132: 115–130.
- Sunnucks P, et al. 2000. SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol Ecol.* 9(11): 1699–1710.
- Tabachnick WJ. 1996. *Culicoides variipennis* and bluetongue-virus epidemiology in the United States. *Annu Rev Entomol.* 41: 23–43.
- Tay WT, Elfekih S, Court L, Gordon KH, De Barro PJ. 2016. Complete mitochondrial DNA genome of *Bemisia tabaci* cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). *Mitochondrial DNA A DNA Mapp Seq Anal.* 27(2): 972–973.
- Tay WT, et al. 2017. Novel molecular approach to define pest species status and tritrophic interactions from historical *Bemisia* specimens. *Sci Rep.* 7(1): 429.
- Tay WT, Evans GA, Boykin LM, De Barro PJ. 2012. Will the real *Bemisia tabaci* please stand up?. *PLoS One* 7(11): e50550.
- Tay WT, Kerr PJ, Jermin LS. 2016b. Population genetic structure and potential incursion pathways of the bluetongue virus vector *Culicoides brevitarsis* (Diptera: Ceratopogonidae) in Australia. *PLoS One* 11(1): e0146699.
- Thao ML, Baumann L, Baumann P. 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol Biol.* 4: 25.
- Tsagkarakou A, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID. 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull Entomol Res.* 97(1): 29–40.
- Ueda S, Kitamura T, Kijima K, Honda KI, Kanmiya K. 2009. Distribution and molecular characterization of distinct Asian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Japan. *J Appl Entomol.* 133(5): 355–366.
- Wang H-L, et al. 2013. The characteristics and expression profiles of the mitochondrial genome for the Mediterranean species of the *Bemisia tabaci* complex. *BMC Genomics* 14: 401.

Associate editor: Liliana Milani