The trouble with MEAM2: Implications of pseudogenes on species delimitation in the globally invasive *Bemisia tabaci* (Hemiptera: Aleyrodidae) cryptic species complex

Tay WT¹, Elfekih S¹, Court LN², Gordon KHJ¹, Delatte H², De Barro PJ³.

1. CSIRO, Black Mountain Science & Innovation Park, Clunies Ross Street, ACT 2601, Australia
2. CIRAD, UMR PVBMT, 97410 Saint-Pierre, La Réunion, France
3. CSIRO, Ecosciences Precinct, GPO Box 2583, Brisbane QLD 4001, Australia

Abstract

Molecular species identification using sub-optimal PCR primers can over-estimate species diversity due to co-amplification of nuclear mitochondrial (NUMT) DNA/pseudogenes. For the agriculturally important whitefly *Bemisia tabaci* cryptic pest species complex, species identification depends primarily on characterisation 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 1 (MEAM1) species, Middle East Asia Minor 2 (MEAM2), and the Indian Ocean (IO) species. Initially identified from the Indian Ocean island of Réunion, 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 characterise the MEAM2 mitogenome, we instead characterised 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, while presence of internal stop codons, loss of expected stop codons and poor primer annealing sites, all suggested MEAM2 as a pseudogene artefact and so not a real species.

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

**Correspondence:** E-mail: *weetek.tay@csiro.au* Tel: +61-2-6246 4286

© The Author(s) 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
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 analysed. 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 rRNA and 28S rRNA gene regions, e.g., Jorger and Schrodl 2013; microsatellite DNA markers, e.g., Cheng, et al. 2013).

One significant challenge faced in the delimitation of otherwise indistinguishable species using mtDNA COI datasets 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 efficacies (Lobo, et al. 2013; Moulton, et al. 2010; 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 (Bensasson, et al. 2001; Sunnucks, et al. 2000). Numerous methodologies are available to assist with the identification of NUMTs (reviewed in Bensasson, et al. 2001) and include identifying nonfunctionality 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, e.g., cassava brown streak virus (Maruthi, et al. 2005), tomato yellow leaf curl virus (Ghanim, et al. 1998), blue tongue virus (De Liberato, et al. 2005; Tabachnick 1996) and parasites, e.g. 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. 2016c). 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. 2016c).
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 (Lee, et al. 2013; Boykin, 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 2005), 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 characterisations 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, et al. 2004; Tay, et al. 2016; Tay, et al. 2017; Wang, et al. 2013). In this study, we characterised 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 characterisation 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 characterised via the high throughput sequencing method. We assessed and discussed the impact of NUMT on phylogenetic inferences on the cryptic *B. tabaci* species complex.

**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 characterised 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 analysed via NGS from the same collection was identified as belonging to a different member of the complex, AUS (657bp 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 (657bp mtDNA COI partial gene matched 100% sequence identity of a Madagascan IO (AJ550171)) and MED (partial mtDNA COI gene (657bp) 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 behavioural, morphological, and genetic differences (e.g., Bellows, et al. 1994; Perring, et al. 1992; Perring, et al. 1993) and was subsequently named *B. argentifolii* (Bellows, et al. 1994). Thao, et al. (2004) provided partial regions (i.e., Cyt b-COIII, 4,796bp; 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, while 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 2.15-28.96 million paired-end (PE) sequences (Table 1), from which 10,738-131,328 PE sequences were assembled 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 (Figs. 1A and 1B). 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 (Figs. 1A and 1B). Corresponding SNP frequencies across DNA fragments generated from high throughput sequencing, and that potentially represented NUMTs within the 657bp mtDNA COI partial gene region, were detected at very low frequencies (Suppl. Table 1a), 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 (Suppl. Tables 1b, 2b, 3b, 4b, 5b, 6b). For MEAM2 when compared with MED and IO, there were no particular SNP frequency patterns (Suppl. Tables 1c,d; 2c,d; 3c,d; 4c,d; 5c,d; 6c,d). 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 (Suppl. Tables 1a, 2a, 3a, 4a, 5a, 6a). Characterisation 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 artefacts 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, 1632 bp mRNA linear INV 09-Nov-2016; accessed 05-Jan-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 Hazkan-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 artefact. With increasing molecular characterisation 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 Peru individuals initially identified as MEAM2 based on partial mtDNA COI gene using sub-optimal primers, high throughput sequencing have also resulted in MEAM1 mitogenomes being assembled instead. This exercise highlighted the importance
of analysing an adequate number of individuals from a collection and the impact sub-optimal PCR primers can have on estimating species composition. These included misidentification of species composition complexity at the population level, and minimising valuable resources being misdirected to monitor for incursion of non-existent 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; Karut, et al. 2015; Ueda, et al. 2009) 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 (Chu, et al. 2011; Frohlich, et al. 1999; Simon, et al. 1994; Tsagkarakou, et al. 2007) and we suspect, factors such as reduced annealing site specificity (Suppl. Table 7) are contributing to the co-amplification 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 (Suppl. Table 7). 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; Tay, et al. 2017; this study) and will significantly contribute to development of B. tabaci species-specific primers, although standardisation 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 17-March-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.

**Material and Method**

*Bemisia tabaci* samples, gDNA extraction, PCR and NGS

Five individuals of *Bemisia* whiteflies from a single Peruvian population, collected on 14-August 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 analysed 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> (accessed 01-Nov-2016). 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 25µL 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, 1.5ng samples of gDNA were tagmented (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 28ul 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 4nM. The Nextera XT gDNA libraries were pooled, diluted to a final concentration of 11pM (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, et al. 2013) prior to manual re-adjustment within Geneious 8.1.9 to identify potential stop codons in all coding sequences (KY951447, KY951448, KY951449, KY951450, KY951451, KY951452). Assembled draft mitogenomes were re-confirmed for species identity by Blastn searches of the partial (657bp) 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; Karut, et al. 2015; Ueda, et al. 2009); 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, IO) known to be present in countries that have also reported MEAM2 (e.g., Reunion, Turkey, Japan, Peru, 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.

Acknowledgements

Chris Hardy, Tom Walsh and William James (CSIRO) provided valuable discussion during the course of this study. SE was supported by a CSIRO OCE Post Doctoral Fellowship (R-4546-1). WTT and SE 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: WTT, SE, LNC, HD, KHJG, PJDB. Laboratory work: WTT, LC, SE. Data analysis: WTT, SE, Manuscript preparation: WTT, SE, LNC, KHJG, HD, PJDB. All authors have read and agreed to the manuscript.

References


**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). The 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).

<table>
<thead>
<tr>
<th>Species</th>
<th>Total PE-seq</th>
<th>MTG PE-seq</th>
<th>Average COI coverage (± s.d.)</th>
<th>mitogenome lengths†</th>
<th>GenBank</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAM1</td>
<td>6,514,260</td>
<td>13,986</td>
<td>166.9 ± 16.4 s.d.</td>
<td>15,666</td>
<td>KY951450</td>
</tr>
<tr>
<td>MEAM1</td>
<td>28,964,958</td>
<td>131,328</td>
<td>990.3 ± 163.4 s.d.</td>
<td>15,531</td>
<td>KY951449</td>
</tr>
<tr>
<td>MED</td>
<td>6,663,490</td>
<td>12,176</td>
<td>84.0 ± 21.0 s.d.</td>
<td>15,526</td>
<td>KY951452</td>
</tr>
<tr>
<td>IO</td>
<td>2,157,716</td>
<td>11,380</td>
<td>126.7 ± 28.5 s.d.</td>
<td>15,631</td>
<td>KY951447</td>
</tr>
<tr>
<td>AUS</td>
<td>4,981,182</td>
<td>15,438</td>
<td>173.9 ± 25.3 s.d.</td>
<td>15,686</td>
<td>KY951451</td>
</tr>
<tr>
<td>MED</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>15,632</td>
<td>JQ906700</td>
</tr>
<tr>
<td>ASIA I</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>15,210</td>
<td>KJ778614</td>
</tr>
<tr>
<td>ASIA II-7</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>15,515</td>
<td>KX714967</td>
</tr>
</tbody>
</table>

† Mitogenome 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.
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 904-906) 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.