Diversity and Distribution of Cryptic Species of the
Bemisia tabaci (Hemiptera: Aleyrodidae) complex in Pakistan

Mariyam Masood,1,2 Imran Amin,1 Ishtiaq Hassan,1,2 Shahid Mansoor,1
Judith K. Brown,4 and Rob W. Briddon1,5

1Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad 38000,
Pakistan, 2Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan, 3Present address: Centre for
Human Genetics, Hazara University, Mansehra, Khyber Pakhtunkhwa, Pakistan, 4School of Plant Sciences, University of Arizona,
Tucson, AZ 85721, and 5Corresponding author, e-mail: rob.briddon@gmail.com

Received 12 January 2017; Editorial decision 16 July 2017

Abstract

Bemisia tabaci (Gennadius; Hemiptera: Aleyrodidae) is considered to be a cryptic (sibling) species complex, the
members of which exhibit morphological invariability while being genetically and behaviorally distinct. Members
of the complex are agricultural pests that cause direct damage by feeding on plants, and indirectly by transmitting
viruses that cause diseases leading to reduced crop yield and quality. In Pakistan, cotton leaf curl disease, caused
by multiple begomovirus species, is the most economically important viral disease of cotton. In the study outlined here,
the diversity and geographic distribution of B. tabaci cryptic species was investigated by analyzing a taxonomically
informative fragment of the mitochondrial cytochrome c oxidase 1 gene (mtCOI-3). The mtCOI-3 sequence was
determined for 285 adult whiteflies and found to represent six cryptic species, the most numerous being Asia II-1
and Middle East Asia Minor 1 (MEAM-1), the later also referred to as the B-biotype, which was previously thought to
be confined to Sindh province but herein, was also found to be present in the Punjab province. The endemic Asia I
was restricted to Sindh province, while an individual in the Asia II-8 was identified in Pakistan for the first time. Also
for the first time, samples were collected from northwestern Pakistan and Asia II-1 was identified. Results indicate
that in Pakistan the overall diversity of B. tabaci cryptic species is high and, based on comparisons with findings
from previous studies, the distribution is dynamic.

Key words: Bemisia tabaci, whitefly, mitochondrial cytochrome c oxidase I, phylogenetic analysis, cotton

The whitefly Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae) is a phloem-feeding insect that has become a prominent problem in
agricultural crops, particularly in subtropical-mild climate locales, worldwide. This usually polyphagous whitefly affects plants directly
by causing feeding damage, and indirectly by secreting honeydew that promotes fungal growth (sooty mold) on leaves and fruit, as
well as contaminating the lint of cotton plants (Byrne and Bellows 1991). In addition, B. tabaci transmits viruses of at least five
plant virus genera (Navas-Castillo et al. 2011), among which the most important are the single-stranded DNA viruses in the genus
Begomovirus (family Geminiviridae) (Brown et al. 2015).

B. tabaci has become recognized as a complex of cryptic species (see refs in Brown 2010), comprising numerous morphologically
indistinguishable, at least partially reproductively isolated species (Dinsdale et al. 2010, Alemandri et al. 2012, Firdaus et al. 2013,
Hadjistylli et al. 2016). Because of this, molecular sequencing of the mitochondrial cytochrome c oxidase I gene (mtCOI) and phylogen-
etic analysis has been used to discriminate members of the complex (Brown et al. 1995a, Boykin et al. 2007, Brown 2010, Dinsdale et
al. 2010, Xu et al. 2010, De Barro et al. 2011). The genetic diversity of B. tabaci in Pakistan has come under investigation as a result of
the discovery of extensive diversification of begomoviruses in cotton and other crops, and the introduction of exotic B. tabaci and
potentially plant viruses it transmits, owing to the country’s geographic position. In a recent study, at least six cryptic species have
been identified in Pakistan, including a newly discovered variant provisionally referred to as ‘Pakistan’ (Ashfaq et al. 2014).
Cotton leaf curl disease (CLCuD) has been problematic in Pakistan and northwestern India since the early 1990s (reviewed by Sattar et al. 2013). The disease is caused by several distinct begomoviruses in association with a disease-specific betasatellite; cotton leaf curl Multan betasatellite (Brididdon et al. 2001, Mansoor et al. 2003, Amin et al. 2006, Amrao et al. 2010a). Presently, there are no commercial tetraploid cotton varieties with resistance to CLCuD.

The objective of the study reported here was to explore the diversity of B. tabaci throughout Pakistan based on phylogenetic analysis of the 3′-fragment of the mtCOI gene (mtCOI-3′) gene sequence. A comparison of the results with findings from earlier studies points to greater than expected diversity of B. tabaci, and the potential for recent shifting in the geographic distribution of some cryptic species in Pakistan.

Materials and Methods

Sample Collection

Sampling of adult B. tabaci whiteflies followed the protocol of Ahmed et al. (2011). Adult B. tabaci whiteflies were collected from cotton and several other host plant species at 125 different locations throughout Pakistan, during 2012-2014. Samples were collected using an aspirator and were preserved in 85% ethanol until use. The geographic coordinates of the sampling site were obtained using a handheld GPS device (eTrex 10; Garmin, Schaffhausen, Switzerland).

DNA Isolation and PCR Amplification

Total DNA was isolated from single whiteflies using a Fast Tissue-to-PCR Kit (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer. Amplification of the mtCOI-3′ fragment (780 bp) was carried out using the primer pair C1J2195/TL2N3014 (TTGATTTTTTGGTCATCCAGAAGT/ TCCAATGCACTAATCTGCCATATTA) and the cycling parameters described for B. tabaci (Simon et al. 1994, Frohlich et al. 1999) and DreamTaq Green PCR Master Mix (Thermo Fisher Scientific). The polymerase chain reaction (PCR) cycling parameters were one denaturation cycle of 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min, followed by a final extension of 72°C for 7 min.

Cloning, Sequencing, and Sequence Analysis

The PCR products were ligated in the pTZ57R/T plasmid vector (Thermo Fisher Scientific). The sequence of each cloned insert was determined by bi-directional, automated Sanger dideoxy chain-termination sequencing (Genomics and Bioinformatics Research Unit, USDA-ARS, Stoneville, MS) using a capillary ABI 3730XL sequencer (Thermo Fisher Scientific). The DNA sequences were assembled, aligned, and edited using Lasergene software (DNASTAR, Madison, WI). Sequence alignment and pairwise distance analyses were performed using MEGA6 (Tamura et al. 2013). The model of evolution was selected [based on a consensus of the Bayesian Information Criterion (BIC), the corrected Akaike Information Criterion (AICc), and the maximum likelihood parameters] using MEGA6 (Tamura et al. 2013). The Hasegawa-Kishnio-Yano (HKY85) model, with gamma distributed rate of variation among sites, was identified as the best fit model of evolution.

Results

Phylogenetic Analysis of mtCOI-3′ Sequences

The mtCOI-3′ sequence (780 bp) for 285 individual whiteflies from field collection sites was determined bi-directionally for cloned PCR amplicons (Table 1, Suppl Table 1 [online only]). The manually edited final sequences were obtained and are available in the nucleotide sequence databases under the accession numbers given in Suppl Table 1 [online only].

The sequences were trimmed to remove primer sequences and aligned with 550 mtCOI-3′ sequences of 658 bp in size, available in a curated global B. tabaci database (http://doi.org/10.4225/08/50EB54B6F10402) and used to classify them as hypothetical cryptic species, as proposed by Boykin and De Barro (2014). Using this approach, six cryptic species of B. tabaci were identified among the samples, including those that grouped with Asia I, Asia II-1, Asia II-5, Asia II-7, Asia II-8 and Middle East Asia Minor 1 (MEAM-1) (also referred to as the B-biotype; Brown et al., 2010; Table 1). Phylogenetic analysis of representative sequences from Pakistan aligned with examples of the previously identified 36 B. tabaci variants/cryptic species revealed that among the six species, including those endemic to Asia, was an exotic species MEAM-1 (Fig. 1; Boykin et al. 2007, Brown 2010, Boykin and De Barro 2014). A separate analysis containing all of the sequences determined herein showed that the Asia II and MEAM-1 cryptic species groups, respectively, were represented by n = 235 and 38 collections (Table 1; Suppl Table 1 [online only]).

Pairwise distances among sequences range from 0.0 to 19.4%. The largest intraspecific nucleotide variation for mtCOI-3′ sequences was observed for Asia II-1 at 3.3%, followed by MEAM-1, at 1.1%, and then Asia I, at 0.2% (Table 1). Based on a hypothetical 3.5% proposed cutoff for differentiating cryptic B. tabaci species (Dinsdale

<table>
<thead>
<tr>
<th>Cryptic species</th>
<th>Number of insects</th>
<th>Min dist</th>
<th>Mean dist</th>
<th>Max dist</th>
<th>Hosts</th>
<th>Previously reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>285</td>
<td>0.0</td>
<td>6.3</td>
<td>19.4</td>
<td>–</td>
<td>IN, PK, CN, BD</td>
</tr>
<tr>
<td>Asia II-1</td>
<td>233</td>
<td>0.0</td>
<td>0.3</td>
<td>3.3</td>
<td>cotton, tomato, okra, guar, mungbean, sunflower, potato, pumpkin, bean, chilli, bottle gourd, eggplant, castor bean, unidentified weed</td>
<td></td>
</tr>
<tr>
<td>Asia I</td>
<td>9</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
<td>cotton, eggplant, sunflower, tomato</td>
<td>IN, PK, CN, BD, ID CN, IN, PK, IL, IR, ID, US, JP, IT, MA</td>
</tr>
<tr>
<td>MEAM-1</td>
<td>38</td>
<td>0.0</td>
<td>0.3</td>
<td>1.1</td>
<td>cotton, tomato</td>
<td></td>
</tr>
<tr>
<td>Asia II-5</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>cotton</td>
<td>PK, IN</td>
</tr>
<tr>
<td>Asia II-7</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Cotton</td>
<td>PK, IN</td>
</tr>
<tr>
<td>Asia II-8</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>mungbean</td>
<td>IN</td>
</tr>
</tbody>
</table>

Plants from which insects were sampled.

Regions where the cryptic species has previously been identified. Countries are denoted by their standard two letter abbreviations.
et al. 2010) several major phylogeographical groups were resolved. The greatest pairwise sequence divergence for the Pakistan *B. tabaci* mtCOI-3′ sequences was observed for Asia II-1 clade, which diverged from the other clades, at 3.3%. For Asia II-5, Asia II-7 and Asia II-8 clades, the between-clade pairwise distances were 1.8, 2.6, and 1.6% divergence, indicating low differentiation between these groups, and
less than proposed working cryptic species cutoffs, at 3.5–4.0% (Dinsdale et al., 2010, Lee et al., 2013) for hypothetical cryptic species, indicating minimal diversification among them, compared with the Asia II-1 group members.

**Geographic and Host Distribution of *B. tabaci* Cryptic Species in Pakistan**

The geographical distribution within Pakistan of cryptic *B. tabaci* species determined here is summarized in Fig. 2. Results indicated that in the Punjab province, the most common cryptic species was Asia II-1. However, some individuals were identified as the exotic MEAM-1 (B-haplotype) in the southern region that borders Sindh province, whereas, in Sindh province, MEAM-1 was the most common cryptic species, with relatively moderate numbers of Asia II-1 interspersed. For samples analyzed for the first time from the Khyber Pakhtoon Khw province in northern Pakistan, the only cryptic species identified was Asia II-1. Also, three specimens of Asia II-5, and at a low frequency, Asia II-7 and Asia I cryptic species were identified in both the Punjab and Sindh provinces. Additionally, one Asia II-8 specimen was collected near the Punjab border of India.

Of the 285, *B. tabaci* specimens analyzed, 251 were collected on cotton and 36 on other plants species (Supp. Table S1 [online only]). Of the specimens from cotton, 207 were identified as cryptic species Asia II-1 with 176 originating from the Punjab province and 31 from the Sindh province. For MEAM-1 36 specimens came from cotton; 34 from the Sindh province and only 2 from the Punjab province. Keeping in mind that encountering an insect on a plant species does not necessarily indicate that the plant species is a host, the overwhelming numbers of Asia II-1 would suggest that cotton is a major host of this cryptic species. For MEAM-1, of the 38 specimens identified, only two came from plants other than cotton (an unidentified cucurbit species). For the other cryptic species too few individuals were identified to draw any meaningful conclusions concerning host association.

**Discussion**

The genetic diversity and geographical distribution of *B. tabaci* are of great interest because of its importance as the insect vector of begomoviruses (Polston et al. 2014). In Pakistan and western India, the most prominent and widespread virus of cotton is the leaf curl virus complex that causes CLCuD, which causes extensive losses to the cotton crop. Three cryptic *B. tabaci* species groups have previously been identified in Pakistan (Ahmed et al. 2011); two endemic cryptic species groups, Asia I and Asia II-1, and the exotic MEAM-1 (B-haplotype) which has its origins in the Middle East-Mediterranean, North African region (Brown 2010). Recently, Ashfaq et al. (2014) reported six cryptic species, Asia II-1, MEAM-1, Asia I, Asia II-5, Asia II-7, and a previously unreported variant, provisionally referred to as ‘Pakistan’, until further verification is possible. In the study described here, analysis of COI-3′ sequences corroborated the presence of six cryptic species in Pakistan, with the sixth type represented by Asia II-8, instead of the ‘Pakistan’ variant. Whether the recently documented prevalence is the result of the increased interest and study of *B. tabaci* diversity in Pakistan, or is the consequence of real changes in or expansion of *B. tabaci* populations is unclear.

The transmission of begomoviruses by cryptic *B. tabaci* species has been shown to vary in efficiency (see Polston et al. 2014 and references therein). For example, MEAM-1 and Mediterranean (MED, Q haplotypes) were found to transmit TYLCV with equal efficiency, whereas transmission efficiency by Asia II-1 (esterase groups K and P) was about half that of the other two cryptic species (Li et al. 2010). Similarly, MEAM-1 (B-haplotype) transmitted *Papaya leaf curl China*
virus more efficiently than MED, Asia I (esterase groups H and M), and Asia II-7 (Cv; Guo et al. 2015). No similar studies have been reported for transmission of the CLCuD complex by different B. tabaci cryptic species. Nevertheless, the prevalence of Asia I-1 (esterase groups K, P) in Punjab province, where CLCuD is endemic, and the earlier absence of this cryptic species in Sindh province, where CLCuD has been rarely reported, suggests that the predominant vector of CLCuD is has been Asia I-1 (Ahmed et al. 2011). Recently, Asia I-1 has been encountered more frequently in Sindh Province (Ashfaq et al. 2014), and results presented here corroborate this finding. This putative ingress could feasibly occur either from the Punjab provinces of Pakistan or India, or from Gujarat and Rajasthan states in India where Asia I-1 has been documented (Ellango et al. 2015). The recent increase in begomovirus diseases, in particular the increased incidence of CLCuD in cotton in Sindh, may be associated with the subsequent increased prevalence of Asia I-1 (Panwar et al. 2001, Amrao et al. 2010b). Previous studies indicated that only the exotic MEAM-1 was present in Sindh province (Ahmed et al. 2010, 2011). However, because Asia II-1 is a more efficient vector of the viruses causing CLCuD and performs better on cotton in Pakistan (Ahmed et al. 2014), the recent documentation either suggests that the geographic range of Asia II-1 has recently expanded into the Sindh province or that its presence was missed in previous surveys, or both. Possibly of consequence to this scenario is that the GroEL heat shock protein of a B. tabaci endosymbiont, Arsenophonus nasoniae, associated with an Asia II clade B. tabaci type from cotton in India, interacts with the coat protein of one of the CLCuD-associated begomoviruses to putatively facilitate transmission (Rana et al. 2012).

No previous studies have reported B. tabaci from Khyber Pakhtoon Khwa province (KP; previously known as Northwest Frontier Province) in northwestern Pakistan, and herein, only Asia II-1 was found colonizing plants there. These locales are too cold in the winter to allow overwintering of B. tabaci, suggesting that they have been encountered in these areas as yearly immigrants arriving from the warmer southern areas. Cotton is grown in the southern areas of KP, near Dera Ismail Khan, but not in the northern areas where vegetable production dominates, and the insects studied herein were collected. Thus, one possibility is that most or all begomovirus diseases occurring there are the result of annual introductions by dispersing, viruliferous whiteflies when the weather warms sufficiently to support B. tabaci populations, although presently, the specific begomoviral pathogens associated with disease outbreaks there have not been identified.

This is the first report of the Asia II-8 cryptic species in Pakistan, specifically in the Thar region near the border with India. However, it has been identified in India (Ellango et al. 2015). This suggests that Asia II-8 may have entered Pakistan from India, or that it is endemic to this dry region of the sub-Continent which spans the border. However, Ellango et al. (2015) reported Asia II-8 in Kerala, Tamil Nadu, and Karnataka states of southern India, but not in the states immediately adjacent to Pakistan. This suggests that Asia II-8 found in Pakistan is either endemic elsewhere in the country, or that its occurrence in India has shifted since the study in India was conducted. Ellango et al. (2015) identified only Asia II-1 in states bordering Pakistan, with the exception of the Indian Punjab, in which a member of the major Asia I clade was also identified.

As well as India, Pakistan has borders with Afghanistan, Iran, and China. No information is available on B. tabaci present in Afghanistan and the MEAM-1 (B haplotypes) predominate in Iran (Rajaei Shoorcheh et al. 2008). Some of the same cryptic species have been identified in China and Pakistan, including the endemic Asia I, Asia II-1, Asia II-7, and exotic MEAM-1 also occurs there (Hu et al. 2011). Other important cryptic species identified in China have not been identified in Pakistan (Brown 2010). Among the most important of these is MED (Q haplotypes; see Brown 2010 and references therein), which comprises multiple lineages endemic to the Mediterranean region (Hadjistylli et al. 2016), from where it was introduced into China (Chu et al. 2005, 2006).

Earlier studies of B. tabaci diversity, which either did not use COI sequences or did not use the cryptic species nomenclature proposed by Dinsdale et al. (2010), also looked at the identity of B. tabaci populations present in Pakistan, India, and China. Baoli et al. (2007) did not identify the B-haplotype (now MEAM-1) in India but did show it to be present in China. This study additionally showed populations that group in the Asia I and Asia II clades to be present in India, India, and Pakistan, with the Med type (Q-haplotype subgroup) occurring only in China. Studies based on esterase patterns previously reported the presence of the analogous K-biotype (Asia II-1 group) in Pakistan, and the analogous H esterase type (Asia I group) and of I esterase (Asia I) in India (Bedford et al. 1994, Brown et al. 1993b).

Results described here and in previous studies indicate that the distribution pattern of cryptic species in the B. tabaci complex of Pakistan is relatively diverse and dynamic, depending on the region. Based on previous studies, it appears that Asia II-1 cryptic species may have expanded southward over time, from the Punjab into Sindh, whereas, MEAM-1 (B) has spread, following its introduction, northwards into Punjab province. In addition, it is likely that Asia II-8 present in Pakistan originated from India and the results might indicate that whiteflies migrate during summer months into northern areas. Changes in B. tabaci populations have long been known to be associated with the appearance of new begomovirus species and changes in begomovirus populations (Legg et al. 2002, Brown 2007, Jiu et al. 2007). The results highlight the need for additional studies of the interactions between the CLCuD complex and different B. tabaci cryptic species that may lead to differential transmission and the emergence of particular leaf curl viral species/strains, including the potential for differential transmission competency. This knowledge of greater than expected diversity among the B. tabaci complex, including the presence of at least one exotic cryptic species (MEAM-1/B haplotypes), is expected to provide new insights about patterns of begomovirus transmission and spread associated with the dynamic nature of B. tabaci cryptic species in Pakistan.

Acknowledgments

The authors are grateful to Ms. Fanny Liu [USDA Mid-South Area Genomics Laboratory (MSAGL)] for assistance with sequencing and Mr. Jorge Paredes (J.K Brown Lab) for providing assistance with phylogenetic analysis. This report is based upon work supported by the ‘Pak-US cotton productivity enhancement program’ of the International Center for Agricultural Research in the Dry Areas (ICARDA) funded by United States Department of Agriculture (USDA), Agricultural Research Service (ARS), under agreement 58-6402-0-178E. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDA or ICARDA.

Supplementary Data

Supplementary data are available at Journal of Economic Entomology online.

References Cited


