ELSEVIER

Virus Research 71 (2000) 233–260

Virus Research

www.elsevier.com/locate/virusres

Molecular markers for the identification and global tracking of whitefly vector–*Begomovirus* complexes

J.K. Brown *

Department of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

Abstract

Recent unprecedented upsurges in populations of the whitefly *Bemisia tabaci* (Genn.) have drawn much attention to its worldwide importance as an insect pest and as the vector of emergent begomoviruses (Family: *Geminiviridae;* Genus: *Begomovirus*). Several begomoviruses that are considered 'new' and others previously regarded as minor pathogens have been linked to recent epidemics. Recent studies have revealed much variation in begomoviruses, despite the view that DNA-containing viruses do not rapidly accumulate mutations. Also, certain *B. tabaci* 'variants' are known that more effectively or selectively transmit certain begomoviruses and exhibit biotic differences that may influence their spread. Patterns of distribution and dissemination of begomoviruses transmitted by *B. tabaci* are poorly understood because standardized molecular-based tracking methods have not been available. Understanding virus/whitefly vector/host plant interrelationships in the context of emerging problems can be achieved only by linking predicted evolutionary histories with epidemiology using molecular phylogenetic approaches.

Identification and validation of informative molecular sequences are essential initial steps in this process. Genus-wide degenerate polymerase chain reaction (PCR) primers have been developed to amplify and sequence the 'core' region of the coat protein open reading frame (ORF) (V1), permitting 'universal' detection and provisional virus identification by comparisons with described viral genotypes. In subsequent studies reported here, several potentially informative viral ORFs and a non-coding region are explored. Of particular use for expanding diversity studies are group- or virus-specific sequences that can be targeted by utilizing newly available *core CP* sequences, or additional conserved regions around which broad spectrum primers can be designed to target variable sequences in key ORFs or non-coding regions. Prospective markers under exploration were selected with a basis in the most highly conserved viral ORFs, *CP* (V1) and a portion of replication-associated protein (REP) (L1/C1), and a key non-coding sequence that contain sufficient variability and/or virus-specific sequences, and are consequently of potential epidemiological relevance. Because *B. tabaci* occurs as a cryptic species, or species complex, that exhibits biotic polymorphism, yet morphological invariance, traditional morphologically based identification is impossible. An overriding complication to establishing molecular markers for identifying whitefly vector variants is that whitefly sequences in general, have not been available. However, recent work has shown that a partial mitochondria cytochrome oxidase I (mt COI) sequence separates vector variants with a basis in geographical origin, suggesting it is useful for further exploring variability and the phylogenetic history of whiteflies on a large scale. Here, the utility of whitefly mt COI nucleotides (nt) sequences is illustrated for inferring relationships between *B. tabaci* collected from

* Tel.: +1-520-6211402; fax: +1-520-6218839.
E-mail address: jbrown@ag.arizona.edu (J.K. Brown).

0168-1702/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0168-1702(00)00221-5

major world regions. Used collectively, these approaches may permit investigations of the patterns of distribution and dissemination of begomovirus–whitefly vector complexes for the first time. Ultimately, more immediate recognition of exotic viruses and whitefly vectors and early detection of upsurges in vector populations and of emerging viruses will be possible. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Aleyrodidae; *Bemisia tabaci*; Emergent viruses; *Geminiviridae*; Geminiviruses; Insect vector; Invasive species; Invasive species; Molecular epidemiology; Sweetpotato whitefly

1. Introduction

Patterns of distribution and dissemination of begomoviruses (Genus: *Begomovirus*; Family: *Geminiviridae*) that are transmitted by the whitefly vector *Bemisia tabaci* (Genn.) (Homoptera/Hemiptera: Aleyrodidae) are poorly understood, primarily because reliable methods for tracking these viruses and their vector are unavailable. As whitefly-transmitted geminiviruses have become recognized as emergent pathogens in worldwide agro-ecosystems, the need for large-scale epidemiological information has increased. Recent studies have shown that biological differences between vector populations may influence dissemination of begomoviruses in crop and weed species (Bedford et al., 1994; Bird and Maramorosch, 1978; Burban et al., 1992; Brown et al., 1995b; Costa and Brown, 1991; McGrath and Harrison, 1995).

Pervasive evidence has linked new begomoviruses and those previously considered benign, because of their seemingly limited potential for invading crop species (Bird, 1957; Bird and Sanchez, 1971; Costa, 1976; Costa and Russell, 1975; Muniyappa, 1980), to new disease outbreaks in sub-tropical areas (Brown, 1994; Brown and Bird, 1992; Idris and Brown, 1998; Paximadis et al., 1999; Polston et al., 1994) and, more recently, to epidemics in relatively mild, temperate agro-ecosystems (Brown et al., 2000b; Byrne et al., 1990 Idris et al., 2000; Polston and Anderson, 1997). Recent studies have uncovered much variation in begomoviruses, despite the view that DNA-containing viruses do not rapidly accumulate mutations. Evidence suggests that certain begomoviruses, when present in mixtures, can potentially employ pseudo-recombination or reassortment strategies and recombination at certain 'hot spots', such as the 'origin of replication' (Stanley, 1995) or, possibly, other highly conserved viral sequences (Bisaro, 1994; Ilyina and Koonin, 1992). Moreover, it is becoming increasingly clear that begomoviruses exist as a pool of genotypes from which 'new' viruses can emerge with the potential for greater or lowered virulence and/or altered transmission efficiencies or host range phenotypes (Zhou et al., 1997, 1998).

Recent studies have also demonstrated that certain *B. tabaci* vector variants differ in ability to transmit certain begomoviruses and that transmission can be more or less effective (Bedford et al., 1994; Bird, 1957; McGrath and Harrison, 1995). Moreover, biotic variation, including host preference and range, fecundity and dispersal behaviour, among others, is known to influence begomovirus disease epidemiology (Brown et al., 1995b; Burban et al., 1992; Costa et al., 1993). A more complete understanding of specific relationships between begomovirus–whitefly vector–host plant complexes that lead to the emergence or disappearance of begomoviruses can be achieved by linking predicted evolutionary histories and epidemiological information using molecular phylogenetic approaches.

Identification and validation of informative molecular markers is an essential initial step in this process. From such studies, it will become possible to estimate genetic diversity within a virus and its known strains, estimate patterns and rates of mutation for each viral gene or non-coding region, identify causal strains associated with new outbreaks and compare them to previously described viruses. By linking viral strains through phylogenetic and epidemiological information, corresponding plant host ranges can also be identified and associations can be made with specific vector variants. Specifically, it is important to consider (1) biotic features, including pathogenic-

ity, symptom phenotype and virulence, identity of the vector variant and transmission efficiency for relevant viruses, and its (2) predicted evolutionary 'origin' of a species, inferred by biogeographical distribution, and phylogenetic associations revealed by DNA sequence analysis. These combined approaches will facilitate more rapid recognition of newly emerging or exotic begomoviruses and aid in crop improvement efforts for important viral genotypes.

1.1. Begomoviruses as emergent pathogens

Although virus-like diseases associated with whiteflies were reported in crop and indigenous, non-cultivated plants in sub-tropical and relatively mild temperate agro-ecosystems since the early 1900s (Bock, 1982; Brown, 1994; Muniyappa, 1980; Varma, 1963), there have been recent upsurges in the 1970s and 1980s in India, Sudan, and the Western Hemisphere, including Brazil, Central America, Mexico, and the US (Bird and Maramorosch, 1978; Bock, 1982; Brown, 1990, 1994; Brown and Bird, 1992; Costa, 1976; Traboulsi, 1994). Crops typically infected by whitefly-transmitted viruses in the Eastern Hemisphere are cassava, brassicas, tobacco, tomato, and legumes (*Vigna* and *Phaseolus* species) (Brown, 1990; Muniyappa, 1980; Mound, 1983); Western Hemisphere hosts have been bean (*Phaseolus vulgaris*), cotton, soybean, tobacco, and rarely, tomato (Bird and Maramorosch, 1978; Brown, 1994; Brown and Bird, 1992; Brown et al., 1995b; Paximadis et al., 1999; Polston and Anderson, 1997). The aetiologies of the majority of these diseases were generally poorly studied until recently.

In the 1970s, several important breakthroughs occurred and, for the first time, linked plant viruses with transmission by whiteflies as a common phenomenon (Duffus, 1987). The most damaging whitefly-transmitted viruses, now known to be begomoviruses, were associated with a single species of whitefly, *B. tabaci* (Bird and Maramorosch, 1978; Bird and Sanchez, 1971; Bock, 1982; Costa, 1976; Mound, 1983; Muniyappa, 1980; Varma, 1963). By the 1990s, begomovirus-incited epidemics were reported in many world locations. When the exotic, B-type vector, which originated in the Eastern Hemisphere, was introduced and became common in Western Hemisphere sites during 1986–1992, widespread emergence of tomato and pepper-infecting begomoviruses was reported for the first time throughout the Americas. A major driving force has been the ability of the B-type to readily colonize Solanaceous hosts, including eggplant, pepper, tobacco, and tomato, whereas, New World *B. tabaci* did not until quite recently. Regardless of whether upsurges in whitefly vector populations involve indigenous or the exotic B-type, diseases caused by begomoviruses typically occur in irrigated vegetable and cotton agro-ecosystems and affect many crop and wild plant host species (Brown, 1990, 1994; Brown and Bird, 1992; Polston and Anderson, 1997; Traboulsi, 1994).

Initially, whitefly-transmitted geminiviruses were intractable, owing to their general restriction to phloem tissues and recalcitrance to experimental transmission with sap or purified preparations (Goodman, 1981; Harrison, 1985). When observed for the first time by transmission electron microscopy, these viruses were found to have a highly unusual, paired icosahedral capsid morphology (Brown and Bird, 1992; Goodman, 1981). Further, paired icosahedra, or 'geminate' particles, contained one single-stranded, circular DNA molecule, so making geminiviruses the second group of DNA-containing plant viruses and the only group having a positive-stranded, circular ssDNA genome. The first whitefly-transmitted geminiviruses discovered contained two genomic components, A and B (or DNA-1 and -2), encapsidated in separate virions, or particles, whereas, a begomovirus with a monopartite genome, *Tomato yellow leaf curl virus-Is* (TYLCV-Is), was later discovered in Israel by Navot et al. (1991).

In 1978, viruses of what is now known as the family *Geminiviridae* were distinguished as forming a unique group of plant viruses. The group contained two subgroups of Cicadellidae-transmitted geminiviruses (subgroups I and II) and the whitefly-transmitted, or subgroup III geminiviruses. Recent efforts to standardize taxonomy and nomenclature have resulted in establishment

of four genera within the *Geminiviridae*, with all whitefly-transmitted viruses placed within the genus: *Begomovirus*. The name of this genus is derived from the first two letters in the name of the type species, *Be*an *go*lden *mo*saic *virus* (Padidam et al., 1995; Rybicki, 1994, 1998; Mayo and Pringle, 1998).

Begomoviruses have attracted much attention as a novel group of plant viruses owing to their unique ssDNA genome and circular topology, an ability to infect dicotyledonous plants and the ease with which the genomes can be manipulated using molecular approaches. Much has been learned about the specific function(s) of viral-encoded genes and the replication strategy employed. DNA-A and DNA-B components of bipartite begomoviruses are each *c*. 2.6 kilobases (kb) in size and share a non-coding intergenic, or 'common region' (CR) of *c*. 180–200 nucleotides (nt) that is typically identical for cognate components of the same virus species. The CR contains modular *cis*-acting elements of the origin of replication (*ori*) (Fontes et al., 1994) while five open reading frames (ORFs) capable of encoding proteins >10 kiloDaltons (kDA) in size are conserved among the DNA-A component of begomoviruses. The coat protein (*CP*) is encoded by the ORF (AV1) and is the most highly conserved gene of the begomoviruses. The replication-associated protein (REP) encoded by the AC1 ORF initiates viral DNA replication and specificity of replication is mediated through interactions of REP with *cis*-acting elements of the *ori* (Bisaro, 1996; Gladfelter et al., 1997; Jupin et al., 1995; Laufs et al., 1995a,b; Lazarowitz et al., 1992; Orozco et al., 1997, 1998). DNA-B encodes two polypeptides, BV1 and BC1, both of which are essential for systemic movement and have been shown to influence host range (Ingham et al., 1995; Jeffrey et al., 1996; Lazarowitz, 1991; Pascal et al., 1994; Sanderfoot and Lazarowitz, 1995, 1996; Noueiry et al., 1994).

Monopartite begomoviruses have a genome of *c*. 2.7–2.8 kb in size and all viral genes and a non-coding, intergenic region are found on a single viral component. Monopartite viral *cis*-acting elements involved in replication are located in the non-coding intergenic region (*c*.200 nt, analogous to CR sequences in bipartite genomes) and certain gene products such as REP are identical in function to the analogous ORFs of their bipartite relatives. However, monopartite viruses differ from their bipartite counterparts in certain respects, including ORFs involved in nuclear transport and long distance movement functions which are ascribed to *CP* and other viral-encoded proteins whose functions are not yet fully understood (Bisaro, 1996; Lazarowitz, 1992).

1.2. Prospective molecular marker strategies for begomoviruses

The recognition of diversity within begomoviruses has created a need for standard approaches to identify, distinguish and track well known and new, previously uncharacterized viruses. For plant viruses, serology has long been the method of choice, but virus-specific antisera for begomoviruses are not available due to the extremely conserved coat protein sequences across species in the genus. Consequently, DNA-based diagnostic approaches, including polymerase chain reaction (PCR) amplification and DNA sequencing of viral sequences have supplanted serology for detection, identification, and classification of this group. Nucleotide or amino acid (AA) sequence alignments are presented conveniently as cladograms to illustrate predicted relationships among taxa and sequence distances can be calculated to estimate divergence.

Sequence-based predictions are known to yield certain incongruences in inferred begomovirus relationships, depending on the specific viral ORF or sequence considered. The extent of conservation and variability and the extent to which sequences can vary without disrupting functionality may be revealed in part by the specific functional roles of a partial ORF or gene sequence. Consequently, specific regions of the genome exhibiting conserved nt or AA sequences may provide insights into variability and, thereby, to begomovirus relationships at higher orders, i.e. family, genus, and species, whereas, lower order species- or strain-differences may be revealed by more variable sequences that can vary without hindering function (Padidam et al., 1995; Rybicki, 1994, 1998). Moreover, lineage-specific sequences may

reveal genealogical insight within or outside the *Geminiviridae* (Ilyina and Koonin, 1992; Lazarowitz, 1992; Rigden et al., 1996; Rybicki, 1994).

Because substantial information about begomovirus genome organization and gene function is now available, as are numerous genomic and coat protein (*CP*) gene sequences, it is possible to scan partial or entire genomic sequences and develop hypotheses regarding the prospective utility of a particular ORF or non-coding region for inferring phylogenetic relationships. For example, sequences associated with functional domains of viral polypeptides, nt sequence elements that exhibit a high degree of conservation within the genus or a distinct species, or sequences that are highly variable within a species or strain have been explored (Arguello-Astorga et al., 1994; Brown et al., submitted for publication, Padidam et al., 1995; Rybicki, 1994).

The only begomovirus sequence formally approved by the ICTV for indicating begomovirus identity is the highly conserved *CP* gene (Mayo and Pringle, 1998) and its translation product. This basis for the high degree of *CP* sequence conservation is due to its multi-functional role in the viral infection cycle. The *CP* is involved in whitefly-mediated transmission, virion capsid assembly and directly or indirectly, in viral movement within the host (Harrison, 1985; Sanderfoot and Lazarowitz, 1996). This protein exhibits moderate variation within the N-terminal $\sim 55-60$ AA, while the central portion of the ORF contains both variable and conserved regions and C-terminal AA are nearly identical for all viruses. An arbitrary value of 90% nt sequence identity has been suggested as a guideline for predicting the identity of a distinct begomoviral species ($<$ 90%), or viral strain ($>90\%$) (Padidam et al., 1995). There is clear evidence that changes to one or several key AA can lead to altered symptom or host range phenotypes (Brown, 2000; Lazarowitz, 1991; Petty et al., 1995).

A limitation to an initial reliance on the *CP* is that this ORF is not readily accessible using a degenerate PCR-based approach for most begomoviruses. Nor does it yield a PCR product of a manageable length for DNA sequencing in a single reaction, both essential for wide-scale studies of diverse genotypes (Brown et al., submitted for publication, Wyatt and Brown, 1996). To circumvent this problem, the *CP* was examined to identify an internal sequence that would be both available using genus-specific or 'universal' PCR primers and informative for initially establishing provisional virus identification. Further, an ideal sequence should be sufficiently short to enable its sequencing in a single reaction and be useful for the subsequent design of virus-specific primers pending a need for additional scrutiny. A segment that meets these criteria was located in the approximate middle, or 'core', of the *CP* gene. This fragment extends between nt 324 and nt 889 (originally 514 and 1048), and is referred to as the '*core CP*' sequence (Brown et al., submitted for publication). The *core CP* sequence contains a region consisting of highly variable and conserved nt stretches, is defined by two highly conserved flanking regions for all members of the genus, can be amplified by PCR using a single pair of degenerate primers (Wyatt and Brown, 1996), and yields an amplicon of 576–579 nt in length whose sequence can be obtained in a single reaction by most automated sequencers (Brown et al., submitted for publication). Use of the *core CP* sequence has allowed us to establish the provisional identity of numerous field isolates from a variety of crops and weed species using multiple sequence alignment or BLAST algorithms (Brown et al., submitted for publication). In addition, this approach permitted the discovery of several (provisional) 'new' begomoviruses. These were subjected to further investigation using the available *core CP* sequence for the design of non-degenerate PCR primers with which to PCR-amplify and obtain the complete *CP* sequence (Brown et al., 2000b; Idris and Brown, 1998; Idris et al., 1999) recommended by the ICTV for establishing the provisional identification of begomovirus species (Mayo and Pringle, 1998).

A second informative sequence is present within the 200 N-terminal AA of the AL1/C1 ORF, which encodes the REP-associated protein (Bisaro, 1996; Laufs et al., 1995a,b) and regions of this polypeptide have recently been shown to be virus- and possibly group-specific. One virus-specific domain interacts with the *cis*-acting REP-

binding elements within the *ori* during initiation of replication, while less species-specific functional domains also occur in this region of the polypeptide, including an ATP-binding domain and nicking-closing functions (Hanley-Bowdoin et al., 1999). Computer-based predictions for *Tomato golden mosaic virus* (TGMV) AL1 indicated that two sets of alpha helices (1 and 2) separated by several AA residues in the first 160 AA of the polypeptide (Orozco et al., 1997) and studies with *Bean golden mosaic virus* (BGMV) revealed an ATP-binding domain downstream from the N-terminal 220 AA (Hansen et al., 1995). The N-terminal 200 AA of this multi-functional gene contains sequences involved in oligomerization of AL1 protein (forming a multimer), DNA-binding and DNA cleavage and joining functions (Bisaro, 1996; Eagle et al., 1994; Heyraud-Nitschke et al., 1998; Laufs et al., 1995a,b). Hence, sequences in the N-terminus region are essential for recognition and/or binding to the virus-specific sequence elements in the common/intergenic region to initiate replication and are also thought to govern certain virus–host interactions. Although the AL1/L1 ORF itself is less conserved overall than the *CP*, the N-terminal 200 AA sequences of this polypeptide may yield an interesting view of begomoviral genealogy and may provide insights into potential recombination events (Bisaro, 1994; Stanley, 1995) denoted by phylogenetic predictions incongruent with those of *CP* trees, as well as clues about genetic drift or modular protein evolution (Fuchs and Buta, 1997; Hanley-Bowdoin et al., 1999).

A third potentially informative virus-specific or lineage-specific region is a direct and/or inverted repeat (or nearly so) which is a non-coding sequence of about 20–40 nt in the CR/IR, just leftward of the nonanucleotide sequence of all begomoviruses. These 'iterated' or repeating sequences were originally postulated to have a virus-specific function owing to their conserved 'signature' in lineages of closely related strains or viral species (Arguello-Astorga et al., 1994). This sequence element was shown recently to play a direct role in replication by its ability to bind REP during initiation of viral replication (Bisaro, 1996; Hanley-Bowdoin et al., 1999). In bipartite begomoviruses, this element is characterized by two 4–5 nt directly-repeated sequences with 2–4 'spacer' nt and it has been shown to participate in REP-binding in a virus-specific manner (Behjatnia et al., 1998; Orozco et al., 1997, 1998; Hanley-Bowdoin et al., 1999). In monopartite begomoviral genomes this element typically consists of a longer stretch of nt that is often difficult to discern and contains direct and inverted repeats, plus 'spacer' sequences.

Although a non-coding region, this element and its putative 'spacer' sequences are clearly species- and at times, lineage-specific (Table 3). Consequently, when combined with a systematic analysis of *CP* and REP markers, this element may provide information that accurately reflects viral variability owing to replication-related selection factors and so may indicate the extent and type of changes that are required for the evolution of virus–virus and virus–host-specific replication determinants at the species level. Consequently, these elements have been investigated as potential informative indicators of viral genealogy that are traceable through strains of extant viral species (Brown et al., 2000a; Paximadis et al., 1999) and between closely related but distinct begomoviral lineages (Arguello-Astorga et al., 1994 Brown et al., 1999; Brown et al., 2000b; Idris and Brown, 2000).

Further, directly repeat sequence elements in the IR/CR of begomoviruses are thought to have a prospective function in the formation of a 'new virus' through intermolecular recombination events that occur during mixed infections. There are several examples in which this prospect has been tested empirically and infection by reassortants (pseudo-recombinations) that shared identical or similar REP-binding sites has been demonstrated, whereas species from more highly divergent lineages are incompatible, although they have a common host (Hill et al., 1998; Hou and Gilbertson, 1996). Interestingly, the progeny B component resulting from a reassortant in tomato comprising *Tomato mottle virus* (ToMoV) and *Bean dwarf mosaic virus* (BDMV) components (two viruses not viewed as closely related based on an overall comparison of their genomic sequences), apparently acquired a portion of the CR

sequence that permitted its most expedient replication, i.e., of the viral component encoding the REP protein (Hou and Gilbertson, 1996).

Approximately 10–30 nt rightward of this virus-specific element is the geminiviral consensus nonanucleotide, TAATATT/AC, containing the nick site (/) in the *ori* (Bisaro, 1996; Lazarowitz et al., 1992). The latter is considered a likely 'hot spot' for recombination (Stanley, 1995). Recombination within the consensus nonanucleotide, accompanied by non-cognate viral, or even foreign, nt could result in the replacement of one or both direct repeats within the element, resulting in a chimeric, and hence, unique REP-binding site (Bisaro, 1994; Hou and Gilbertson, 1996). If new viable, sustainable viruses emerge commonly in this way, initially, a routine determination of the sequence within this diagnostically important non-coding region would allow for predictions concerning the feasibility of prospective reassortant and recombinant events among extant begomoviruses.

1.3. The whitefly vector of begomoviruses

All begomoviruses are transmitted by a single species of whitefly, *B. tabaci*, yet strong evidence for biotic variation exists, despite the absence of useful morphological characters that permit their distinction (Gill, 1992a). Early reports suggested that biological races of *B. tabaci* could either limit or facilitate whitefly-mediated transmission of certain begomoviruses (Bird, 1957; Bird and Maramorosch, 1978; Costa and Russell, 1975).

Specific vector traits may either directly or indirectly influence the emergence of 'known' begomoviruses and feasibly could also control virus host range, as for *Jatropha mosaic virus* (JMV) which is transmitted by the monophagous Jatropha biotype (Bird, 1957, Brown et al. 1993). In fact, recognition of 'races' of *B. tabaci* that exhibited different host ranges or host preferences (Bird, 1957; Bird and Sanchez, 1971; Bird and Maramorosch, 1978; Brown et al., 1995a) supplied the first evidence of a prominent biotic feature that could drive selectivity of transmission and, thereby, vector-mediated dissemination of begomoviruses. In Puerto Rico, the highly polyphagous 'Sida race' of *B. tabaci* (which feeds and reproduces on many species of *Sida*) was identified as the vector of over 30 begomovirus diseases of weed and crop species (Bird and Maramorosch, 1978; Bird and Sanchez, 1971; Brown and Bird, 1992). In contrast, the restriction of JMV to *Jatropha gossypifolia* (L.) (and its close relatives) occurred because the monophagous Jatropha race was unable to utilize other species as hosts (Bird, 1957; Brown et al., 1995a; Frohlich et al., 1999).

Selective begomovirus dissemination in cassava worldwide may be explained because *B. tabaci* could not colonize cassava in its centre of diversity in South America (Costa and Russell, 1975), yet since the first introduction of cassava to Africa during the 1500s–1700s, it has become a widely grown staple crop which is readily colonized by *B. tabaci*, an important vector of several cassava-infecting begomoviruses (Bock, 1982; Burban et al., 1992; Zhou et al., 1997). Likewise, *B. tabaci* colonizes cassava in India and is the vector of *Indian cassava mosaic virus* (ICMV) (Muniyappa, 1980; Varma, 1963) which is also of Old World origin.

With widespread and frequent upsurges in *B. tabaci* populations in agro-ecosystems throughout the sub-tropics (Brown and Bird, 1992; Brown, 1990, 1994; Byrne et al., 1990; Cock, 1993; Mound and Halsey, 1978; Traboulsi, 1994), much attention has been given to exploring more precisely the biotic and genetic differences between isolated *B. tabaci* populations (Burban et al., 1992; Brown et al., 1995a,b; Costa and Brown, 1991; Costa et al., 1993; Frohlich et al., 1996, 1999; De Barro and Driver, 1997; De Barro et al., 2000; Gawel and Bartlett, 1993; Guirao et al., 1997; Perring et al., 1993; Wool et al., 1993) and, further, to understanding how such variability influences begomovirus epidemiology. Biotic variation is manifest primarily as differences in host range and/or host specialization, degree of fecundity, variability in virus transmission efficiency and, for the widespread 'type B', the ability to cause phytotoxicity (Bedford et al., 1994; Bird, 1957; Brown et al., 1995a,b; Costa and Brown, 1991; Costa et al., 1993). Clearly, these features and others yet undetermined have great potential to influence or alter the epidemiology of begomoviruses and feasibly can facilitate or undermine opportunities for recombination and reassortant events and, thereby, evolution.

A notable example in which specific biotic features have greatly influenced its unprecedented status as a pest and vector of emerging begomoviruses is provided by the widespread and extremely damaging *B. tabaci* variant known as the 'type B', or silverleaf, whitefly. This Old World variant has been transported inadvertently on ornamental plants throughout much of the world since *c.* 1986 (Brown et al., 1995a,b; Brown et al., 1999; Costa et al., 1993; Gill, 1992b). Associated with the establishment of the B-type variant are emergent begomoviruses such as TYLCV-Is, which originates from the same region as the *B. tabaci* type B. TYLCV-Is was accidentally introduced to the Caribbean in 1992 (Nakhla et al., 1994; Polston et al., 1994) and, more recently, to the southern US (Polston and Anderson, 1997), adding to the extensive list of new begomoviruses described from the Americas and Caribbean. Although it is likely that indigenous *B. tabaci* were competent vectors of TYLCV-Is, the dissemination of TYLCV-Is may have been more restricted if its natural, highly polyphagous vector had not also been present.

The widespread dissemination of the B-type of *B. tabaci* has created much recent interest in the role of *B. tabaci* as a virus vector and pest. However, exhaustive light microscopic (Bedford et al., 1994) and transmission electron microscopic (Rosell et al., 1997) examinations of a suite of biogeographically representative *B. tabaci*, using the morphological characters employed in whitefly identification, revealed no distinctive characters useful for distinguishing variants. This extreme invariance in morphological characters, together with widespread evidence of biotic and genic polymorphism (Bedford et al., 1994; Costa et al., 1994; Brown et al., 1995a,b; Brown et al., submitted for publication; Burban et al., 1992; Coats et al., 1993; Costa and Brown, 1991; Costa et al., 1993; De Barro and Driver, 1997; Gawel and Bartlett, 1993; Perring et al., 1993; McGrath and Harrison, 1995; Wool et al., 1993), has led to the hypothesis that variants of *B. tabaci* comprise a cryptic species or a sibling species complex (Bedford et al., 1994; Brown et al., 1995b; Rosell et al., 1997 Frohlich et al., 1999).

Consequently, development of informative molecular sequences, or 'markers', that can be linked to geographic genotypes (topotypes) and/or biotic phenotypes constitutes the most reliable approach for tracking the distribution and dispersal of *B. tabaci*. Of particular interest has been the basis for vector variation as it pertains to virus emergence and subsequent spread, and to the prospective 'disappearance' of begomoviruses that were once important pathogens. Moreover, relationships between variants revealed by biotic and DNA polymorphisms are relevant to classification of *B. tabaci* variants, in the context of a biological species concept and prospective species boundaries.

Several whitefly nuclear and mitochondria (mt) sequences have been explored for their potential as molecular markers. Nuclear genome sequences explored to date include the small ribosomal subunit rRNA gene (18S rDNA) which varied at a single site for two *B. tabaci* variants (Campbell et al., 1996). Nuclear ITS sequences were utilized recently to make phylogenetic predictions and revealed several distinct groups within *B. tabaci* (De Barro et al., 2000). Best explored are the mt 16S ribosomal gene and the mt cytochrome oxidase I (mt COI) gene sequences (Frohlich et al., 1996, 1999; Kirk et al., 2000; Rey et al., 2000). The mt COI gene has provided the most informative marker to date for resolving groups within *B. tabaci*. These studies revealed groups or clusters of *B. tabaci* with an overriding basis in geographical origin, while work in progress shows some promise in revealing variants that are isolated by host species (author, unpublished data). Here, representative mt COI sequences were selected to illustrate the promising utility of the mt COI sequence for examining the worldwide distribution and relationships between *B. tabaci* variants with relevance to its natural history and begomovirus epidemiology.

1.4. Selection of begomoviral sequences

This overview reports a phylogenetic comparison of four viral sequences (*core CP*, *CP*, REP, REP binding elements) as prospective molecular

markers by which to identify and track begomoviruses. Reference sequences were obtained for a representative suite of well-studied begomoviruses, available in GenBank. Identical taxa were used in all analyses to illustrate the comparative utility of candidate sequences as informative markers. First, the *core CP* was evaluated for its utility in establishing provisional begomovirus identification and subsequent classification to detect new viruses or those already known. Second, taxa relationships inferred by the *CP* tree were compared to the tree topology for a *core CP* tree (Wyatt and Brown, 1996; Brown et al., submitted for publication). Third, *CP* sequence predictions were extended to analyze relationships of over 40 field isolates examined in our laboratory, following *core CP* sequence analysis. Fourth, the N-terminal ~200 AA of the virus-specific REP gene (AL1/C1 ORF) was evaluated as a second possible phylogenetic marker and its congruence with a *CP* AA sequence tree was evaluated. Finally, the recognition of many new begomoviruses in recent years led to a survey of variability and the prospect that family lineages can at times be traced in the non-coding, virus-specific REP-binding element. Results of this cursory 'survey' of begomoviral molecular sequences provide a view of virus relationships based on phylogenetic analysis of coding and non-coding nt and/or AA sequence data and available biogeographic information. Generally, sequence choices were based on knowledge of sequence function and associated sequence conservation or variability, or of virus-specific features.

For *B. tabaci*, selection of a most-informative sequence or sequences have not been established unequivocally. However, much of our recent effort has been directed toward examining a large collection of *B. tabaci* associated with field collections of begomovirus isolates identified using the sequence markers described above. For this purpose, we have employed the mt COI sequence as a molecular marker (Frohlich et al., 1999) in an attempt to unravel global patterns of biotype and topotype distribution and to perhaps elucidate evolutionary relationships between variants of *B. tabaci* in the context of plant host and geographic associations that potentially influence begomovirus epidemiology. Ultimately, our goal is to validate informative molecular sequences for begomoviruses and whitefly vector variants that will form links between their individual and collective evolutionary histories and biotic characters that are relevant to the epidemiology of begomoviruses.

2. Materials and methods

2.1. Representative reference begomoviral sequences from GenBank

Sequences of reference taxa and their GenBank accession numbers [shown in parentheses] for A component/monopartite genomes or the *CP* gene ORF are: *Abutilon mosaic virus*, West Indies (AbMV-WI)[X15983]; AbMV, Hawaii (AbMV-Hi) [U51137]; *African cassava mosaic virus*-Kenya (ACMV-Ke) [J02057]; ACMV-Nigeria (ACMV-Ni) [X68318]; *Ageratum yellow vein virus* (AYVV) [X74516]; *Bean calico mosaic virus* (BCaMV) [AF110189], *Bean dwarf mosaic virus* (BDMV) [M88179]; *Bean golden mosaic virus*-Brazil (BGMV-Br) [M88686]; BGMV-Dominican Republic (BGMV-DR) [L01635]; BGMV-Guatemala (BGMV-Ga) [M91604]; BGMV-Puerto Rico (BGMV-PR) [M10070]; *Cabbage leaf curl virus* (CaLCV) [U65529]; *Chino del tomate virus* (CdTV) [AF226665]; *Cotton leaf curl virus*-Pakistan (CLCuV-Pak) [AJ002458]; *East African cassava mosaic virus*-Tanzanis (EACM-Tan); *Havana tomato virus*-Cuba (HTV-CU) [Y14874]; *Indian cassava mosaic virus* (ICMV) [Z24758]; *Tomato leaf curl virus*, India (ToLCV) [Z48182], *Mung bean yellow mosaic virus* (MYMV) [D14703], *Okra leaf curl virus*-Pakistan (OLCV-Pak) [AJ002459], *Okra yellow vein mosaic virus*-Pakistan (OYVMV-Pak) [AJ002451]; *Pepper huasteco virus* (PHV) [X70418]; *Potato yellow mosaic virus* (PYMV) [D00940], *Sida golden mosaic virus*, Costa Rica (SiGMV) [X99550]; *Sinaloa tomato leaf curl virus* (STLCV) (*CP*) [AF040635]; *South African cassava mosaic virus* (SACMV) [AF01785] (*CP*); *Squash leaf curl virus*-E (SLCV-E) [M38182], *Taino tomato mottle virus* (TToMoV) [AF012300]; *Texas pepper virus*-Tamaulipas (TPV-Tam) [U57457];

Tomato golden mosaic virus (TGMV) [K02029], *Tomato leaf curl virus*-Australia (ToLCV-Au) [S53251]; *Tomato leaf curl virus*-New Delhi (ToLCV-NDe) [U15015]; ToLCV-Taiwan (ToLCV-Tw) [U88692], *Tomato mottle virus* (ToMoV) [L14460], *Tomato yellow leaf curl virus*-Israel, (TYLCV-Is) [X15656]; TYLCV-Is mild (TYLCV-Is mild) [X76319]; TYLCV-Sardinia (TYLCV-Sar) [X61153]; TYLCV-Sicily (TYLCV-Sic) [Z25751]; TYLCV-Spain (TYLCV-Sp) [AJ223505]; and TYLCV-Thailand (TYLCV-Th) [M59838].

2.2. Begomovirus field isolates, PCR, nucleotide sequencing, and phylogenetic analysis

Field samples (Table 1) were collected from symptomatic plants associated with infestations of *B. tabaci* (Brown et al., submitted for publication). Certain samples were obtained from plants inoculated experimentally with viral DNA extracts isolated from field-infected plants, or with full-length infectious viral clones to obtain isolates more widely representative of geography and host species. Leaf tissues were extracted fresh or from frozen or dried samples, depending upon the source and time between collection and extraction. Virus DNA was obtained from fresh or dried leaves by extracting total nucleic acids (0.1 g leaf/300–500 μl buffer) (Doyle and Doyle, 1987; Wyatt and Brown, 1996). Final pellets were redissolved in 300 μl TE buffer (10 mM Tris–1mM EDTA, pH 8.0) and 1–2 μl of total DNA extract was added to a final volume of 25 μl containing a previously standardized PCR master mix (Wyatt and Brown, 1996). *Core CP* fragments were obtained by PCR amplification using degenerate primers V324 and C889 (originally 514 and 1048) that are capable of universal amplification of the majority of begomoviruses (Wyatt and Brown, 1996).

PCR products of the expected size, 576 and 579 bp for Western Hemisphere and Eastern Hemisphere viruses, respectively, were cloned into the pCR.2.1 TA cloning vector (Invitrogen, San Diego, CA 92008) and the nt sequence of inserts was determined. DNA sequences were obtained using an automated DNA sequencer in the Laboratory for Molecular and Systematic Evolution at The University of Arizona, Tucson, AZ. Three to four amplicons per isolate were obtained from a minimum of two cloning experiments and sequences were obtained in both directions using T7 and M13 primers to achieve at least three-fold coverage. Sequences were edited and a consensus sequence was identified for each isolate.

Sequence comparisons were accomplished using the multiple sequence alignment algorithm employed by Clustal (MEGALIGN, DNASTAR, Madison, WI) and subjected to maximum parsimony analysis using PAUP 3.1.1 (Swofford, 1991). A most parsimonious tree was sought using a heuristic search with stepwise addition and the tree-bisection-reconnection branch-swapping options for 200 replicates. Bootstrap values were placed at nodes using a $>60\%$ majority rule option. Pairwise distances were computed by Clustal and PAUP (data not shown).

2.3. Field collections and whitefly taxonomic identification

Adult and/or pupae of *B. tabaci* were collected from crop and weed species in the field by the author, or by laboratory members, colleagues, or commercial contacts from representative locations, worldwide (Table 2) Frohlich et al., 1999; Kirk et al., 2000). When possible, adults and pupae were collected live and placed in 70–95% alcohol until analysis. Pupae, when available, were sent to Mr. Ray Gill, California State Department of Agriculture, Sacramento, CA, for identification. Adults were subjected to molecular analysis except when unavailable and then pupae were removed from leaves and processed as described for adults.

2.4. Polymerase chain reaction, cloning, sequencing, and phylogenetic analysis of whitefly mt COI amplicons

Single whiteflies were prepared for PCR using the lysis procedure of Frohlich et al., (1999). PCR amplification of the mt COI gene fragment and molecular cloning were as described (Frohlich et al., 1999), except cloned amplicons were sequenced

Table 1
Forty two begomovirus field isolates for which the *CP* gene sequence (756–774 nt) was obtained by PCR and compared with reference begomovirus *CP* gene sequences, their *CP* sequence length and geographic and host plant affiliations

Field isolate acronym	*CP* length (nt)	Geographic origin	Field host
abmv Roth/Az	756	UK to AZ, US	*Abutilon striatum*
Abutilon spp Guat	756	Guatemala	*Abutilon* spp.
Ambrosia Bz	756	Campinas, Brazil	Ambrosia spp.
Bean BGMV Hon	756	Honduras	*Phaseolus vulgaris*
Bean PR	756	Puerto Rico, US	*P. vulgaris*
Chile Guat	756	Guatemala	*Capsicum annuum*
CLCV Az	756	Arizona, US	*Gossypium hirsutum*
Cotton 1 Guat	756	Guatemala	*G. hirsutum*
Cotton 1 Texas	756	Texas, US	*G. hirsutum*
Cotton 2 Texas	756	Texas, US	*G. hirsutum*
Cotton DR	756	Dominican Republic	*G. hirstaum*
Cotton Mexacali Mx	756	Sonora, MX	*G. hirsutum*
Cotton sonora Mx	756	Sonora, MX	*G. hirsutum*
E. heterophylla PR	756	Puerto Rico, US	*Euphorbia heterophylla*
Euphorb weed Guat	756	Guatemala	*E. heterophylla*
Hibiscus spp Bz	756	Brazil	*Hibiscus* spp.
JMV PR	756	Puerto Rico, US	*Jatropha gossypifolia*
M. alceifolia PR	756	Puerto Rico, US	*Malva alceifolia*
M. lathyroides PR	756	Puerto Rico, US	*Macroptilium lathyroides*
Malvaceous spp. Guat	756	Guatemala	Malvaceous spp.
Passionvine PR	756	Puerto Rico, US	*Passiflora edulis*
Pepper TPV CPS Mx	756	Chiapas, MX	*Capsicum annuum*
Rhyn. minima PR	756	Puerto Rico	*Rhynchosia minima*
S. rhombifolia PR	756	Puerto Rico	*Sida rhombifolia*
Sida glabra PR	756	Puerto Rico	*S. glabra*
Sida micrantha Bz	756	Puerto Rico	*S. micrantha*
Sida spp Guat	756	Guatemala	*Sida* spp.
Sida spp PR	756	Puerto Rico	*Sida* spp.
Sida spp Tx	756	Texas, US	*Sida* spp.
Sida urens PR	756	Puerto Rico, US	*S. urens*
TBASV Mx	756	Tuxtla, MX	*Nicotiana tabacum*
Tob 11 Mx	756	Tapachula, MX	*N. tabacum*
Tob 46 Mx	756	Nayarit, MX	*N. tabacum*
Tomato 1 Hon	756	Honduras	*Lycopersicon esculentum*
Tomato 2 Hon	756	Honduras	*L. esculentum*
Tomato 3 Hon	756	Honduras	*L. esculentum*
Tomato 4 Hon	756	Honduras	*L. esculentum*
Tomato 5 Hon	756	Honduras	*L esculentum*
Tomato PYMV PR	756	Puerto Rico, US	*L. esculentum*
Tomato Sudan	774	Sudan	*L. esculentum*
Tomato ToMov PR	756	Puerto Rico, US	*L. esculentum*
Watermelon Guat	756	Guatemala	*Citrullus vulgaris*

in forward and reverse directions to obtain a final sequence of 708–720 bp. Primers used to amplify the mt COI gene were C1-J-2195 (5′ ttg att ttt tgg tca tcc aga agt 3′) and L2-N-3014 (5′ tcc aat gca cta atc tgc cat att a 3′) obtained from the UBC Insect Mitochondrial DNA Primer Oligonucleotide Set, compiled by B.J. Crespi and C. Simon (Simon et al., 1994).

PCR products were cloned using a TA-cloning kit by insertion of PCR products into the plasmid pCR2.1 vector (Invitrogen) per manufacturer's instructions. Mitochondria COI clones were se-

Table 2
B. tabaci variants from field collections and whitefly mt COI reference sequences, length of mt COI sequence used in comparisons, and geographic and host plant affiliations (all 'B' type reference collections were identified previously by diagnostic esterase profiles)

B.tabaci collection	COI sequence length (nt)	Geographic origin	Host species
Thailand poinsettia	720	Thailand[a,b]	*Euphorbia pulcherrima*
Thailand Kampcowpea	720	Thailand[a,b]	*Vigna unguiculata*
Spain *lpomoea* M94128	708	Spain[a]	*lpomoea batatas*
SP95 Spain	720	Spain[c]	*Cucumis melo*
SP92 Spain	720	Spain[d]	*Citrullus vulgaris*
Sida Race PR	720	Puerto Rico[e]	*Chamaesyce* spp.
SC-Sudan	720	Sudan[d]	*Gossypium hirsutum*
PC95 Pakistan	720	Pakistan[c]	*G. hirsutum*
PC92 Pakistan	720	Pakistan[d]	*G. hirsutum*
PC91 Pakistan	720	Pakistan[d]	*G. hirsutum*
New Nepal	720	Nepal[d]	*C. vulgaris*
Malaysia Sonchus	720	Malaysia[b]	*Sonchus oleraceae*
Malaysia Malva M94047	720	Malaysia[b]	*Malva parviflora*
JAT Puerto Rico	720	Puerto Rico[e]	*Jatropha gossypifolia*
IW India	720	India[d]	*C. vulgaris*
Italy eggplant M94047	720	Italy[a,b]	*Solanum melongena*
Israel-*lantana* M94115	720	Israel[f,b]	*Lantana camara*
Israel-*Euphorbia* M94125	720	Israel[f,b]	*Euphorbia* spp.
ISC (B) Israel	720	Israel[d]	*G. hirsutum*
HC China	720	China[f]	*G. hirsutum*
GC (B) Guatemala	720	Guatemala[c]	*C. melo*
FC (B) Florida	720	Florida USA[d]	*Solanum nigrum*
CUL Mexico	720	Culiacan Mex[c,g]	*Lycopersicon esculentum*
CC (B) California	720	California USA[d]	*G. hirsutum*
C (B) Texas	720	Texas USA [a,c]	*Brassica oleraceae*
Brawlee-California	724	California USA[h]	*G. hirsutum*
Az (B) Arizona	720	Arizona, USA [c,i]	*E. pulcherrima*
Az (A) Arizona	720	Arizona USA [c,j]	*G. hirsutum*
JW (B) Japan	720	Japan[f]	*C. vulgaris*
Uganda 1	723	Uganda[k]	*Manihot esculenta*
Uganda 2	713	Uganda[k]	*M. esculenta*
Cameroon 1	720	Cameroon[l]	*M. esculenta*
Cameroon 2	720	Cameroon[l]	*M. esculenta*
Cameroon 3	720	Cameroon[l]	*M. esculenta*
Zambia	720	Zambia[l]	*M. esculenta*
Mozambique	720	Mozambique[l]	*M. esculenta*

[a] Kirk et al., 2000.
[b] Collected by A. Kirk and L. Lacey, USDA/ARS-Foreign Exploration.
[c] Arizona collection; J.K. Brown et al.
[d] Bedford et al. (1994); Rosell et al. (1997).
[e] Collected by J. Bird, Univ of Puerto Rico.
[f] Collected by I. Bedford, John Innes Centre, UK.
[g] Costa et al. (1993).
[h] Brown et al. (2000); Frohlich et al. (1999).
[i] Prototype 'B biotype' Costa and Brown (1991).
[j] Prototype 'A biotype' Costa and Brown (1991).
[k] J. Legg and W. Otim-Nape; IITA Cassasva IPM Project, Kampala, Uganda.
[l] S. Berry and C. Rey; Rey et al. (2000).

Table 3
Theoretical REP-binding element (5'–3') in the common/intergenic regions of begomoviruses

Begomovirus taxon	Putative AL1/C1 binding motif			Geographic origin	Partial host range
TYLCV SAR	GGGGG	AATT	GGGGG	Sardina	Tomato
TTMOV	GGGGG	AACT	GGGGG	Cuba/Carib	Tomato
ITOLCV	GGGGG	ACTC	GGGGG	India	Tomato
SACMV	GGGGG	GATC	GGGGG	South Africa	Cassava
CLCUV[a] PAK452; 455[a]	GGGGA	CACT	GGGGA	Pakistan	Cotton; other
HTV; PYMV	GGGGA	ACT	GGGGA	Cuba; Carib & S. Amer.	Tomato; potato
TBLCV IN[a]	GGGGA	CTC	GGGGG	India	Tomato
CDTV/TLCRV	GGGGT	AAA	GGGGA	Mexico, SWUS	Tomato, bean
STLCV	GGGGT	AAT	GGGTT	Mexico, SWUS	Tomato, bean
CLCuV PAK458[a]	GGGAT	ACT	GGGAT	Pakistan	Cotton, other
ABMV; BDMV	GGAGT	ATT	GGAGT	Worldwide; Colombia	*Abutilon*; bean
COT-INF VIRUS1[a,b]	GGAGT	TCT	GGAGT	Guatemala	Cotton
CLCV AZ[b]	GGAGT	CT	GGAGT	SWUS, Mexico	Cotton, bean
TPV	GGAGT	CCT	GGAGT	SWUS, MEX, Carib, C. Amer.	Tomato, pepper
TGMV[c]	GGTAG	TAA	GGTAG	Brazil	Tomato, *N.benthamiana*
TYLCV SP; SIC	GGTAG	ATT	GGTAG	Spain, Sicily	Tomato, bean
ICMV	GGTAC	TCAT	GGTAC	India	Cassava
COT-INF VIRUS2[a,b]	GGTGT	ATT	GGTGT	Guatemala	Cotton
TBASV[a]	GGTGT	ATT	GGTAT	Mexico	Tobacco
PHV	GGTGT	ATT	GGTAG	SWUS, Mexico, C. Amer.	Tomato, pepper
TOLCV TW	GGTGT	ATT	GGGGT	China	Tomato
BGMV BR	GGTGT	AAT	GGTGC	Brazil	Bean
UGV[a]	GGTGT	AAT	GGGGG	Uganda	Cassava
OKLCV PAK459[a]	GGTGT	AAT	GGGGT	Pakistan	Okra, cotton
TYLCV IS, MBYMV	GGTGT	ATC	GGTGT	Israel; Thailand	Tomato, mungbean
TBLCV ZIM[a]	GGTAT	ATC	GGTAC	Zimbabwe	Tobacco
CLCUV PAK448[a]	GGTGT	ATC	GGTGA	Pakistan	Cotton
TOLCV AU	GGTGT	CTG	GGGT	Australia	Tomato
TOLCV ND	GGTGT	CT	GGAGT	India	Tomato
EACMV TAN	GGTGG	TAAC	GGGGT	East Africa	Cassava
COMV[a]	GGTGT	AAC	GGGGT	Africa	Cowpea
BCMOV; SLCV E	GGTGT	CCT	GGTGT	SWUS, Mexico, C. Amer	Bean; squash & bean
CALCV	GGTGT	CTT	GGTGT	Florida US	Cabbage
SIGMV	GAGTA	TTT	GAGTA	C. Amer	*Sida* spp.
TOMOV	GAGTA	TTT	GAGTA	Florida, PR US	Tomato
ACMV KE, NI	GACAA	CATCAACTA	GAGAC	Africa	Cassava
TYLCV TH[a]	GGGGA	AT	GGGTC	Thailand	Tomato
AYVV	GGGGA	AT	GGTAC	SE Asia	Bean, tomato
BGMV CARIB BEAN	GGAG	AC	GGAC	Carib, C. Amer	Bean
CHMV[a]	ATAAA	TC	GGGTC	Africa	Cheramoya

[a] Begomovirus associated with the disease.
[b] J.K. Brown et al., unpublished data.
[c] Mapped to site (Orozco et al., 1997).

quenced at the Laboratory for Systematics and Molecular Evolution, The University of Arizona, Tucson, AZ. PCR products (800 bp) were obtained for a minimum of three whiteflies per collection and at least four clones from each PCR product. Cloned whitefly DNA amplicons were sequenced in both directions using universal primers, as described above. Sequences for each individual whitefly were compiled and a consensus sequence was obtained for each isolate.

Using the same primers a mt COI sequence was also obtained from the outgroups: *Trialeurodes vaporariorum* (West.) the greenhouse whitefly, and *B. berbericola* (Cockerell) the bayberry whitefly.

For ease of alignment, mt COI sequences were edited to 700–720 bp in length. Phylogenetic analysis was accomplished by first aligning sequences with Clustal (MEGALIGN, DNASTAR, Madison, WI), followed by parsimony analysis. Frohlich et al., (1999) examined several approaches by considering different tree-building algorithms that make different evolutionary assumptions, including parsimony, maximum-likelihood and neighbour joining and concluded that the three approaches were in general agreement. Consequently, mt COI sequences were evaluated using distance (Clustal) and maximum parsimony (PAUP, version 3.1.1) (Swofford, 1991) methods as described for virus sequences, except bootstrapping was performed with the heuristic search option for 2000 replicates at the $>60\%$ confidence level. Genetic distances were computed by Clustal and PAUP.

2.5. BLAST analysis and an interactive website for begomovirus and whitefly identification, using GEMINIDETECTive

The BLAST program (Altschul et al., 1997) was downloaded from the internet and installed at the GEMINIDETECT*ive* site, a website (http://gemini.biosci.arizona.edu) that contains information on begomoviruses and provides a comparative tool for establishing provisional begomovirus identification using BLAST (established by J.K. Brown and S.D. Wyatt, ©1997). A database has been compiled at the site containing *core CP* sequences of well-studied begomoviruses and *core CP* sequences obtained from field isolates, strains, or new begomoviral species studied in the laboratory during the past 15 years. The BLAST program is linked to the *core CP* data base, enabling the user to input a *core CP* sequence of an 'unknown' isolate and to receive an 'output' report that provides begomovirus identification, or a 'closest match' with respect to *core CP* sequences available in the database. Users can then link to a 'virus page' to obtain biogeographic information, including worldwide distribution, host range, symptoms in key hosts, and references for a particular begomovirus. A second link enables users to access GenBank directly to carry out BLAST searchers or obtain reference sequences and accompanying information.

3. Results and discussion

3.1. Core region of the coat protein gene as an indicator of preliminary provisional begomovirus identity

The *core CP* region spans seven of the eight most highly conserved AA motifs in the *CP* of begomoviruses. The *core CP* region is flanked by a highly variable 5′-end and the nearly identical 3′-end and is characterized by stretches of nearly identical sequences interspersed with variable bases. These characteristics, along with the presence of highly conserved nt sequences that bracket the core region and serve as sites for universal primer annealing, facilitated the use of *core CP* PCR as a molecular epidemiological tool that simplifies begomovirus detection (Wyatt and Brown, 1996) and identification (Brown et al., submitted for publication). Our focus on the *core CP* is in accordance with the recent ICTV decision that the *CP* gene is an adequate predictor of species versus sub-species standing when this is the only viral sequence available. Obtaining the core fragment first allows one to predict viral identity nearly as accurately as does the complete *CP* sequence and can be accomplished for the majority of begomoviruses using a single pair of PCR primers.

The detection of a begomovirus and ability to 'predict' virus identity by *core CP* sequence comparisons of field isolates with those from well-studied begomoviruses has been demonstrated (Wyatt and Brown, 1996). Here, using diverse

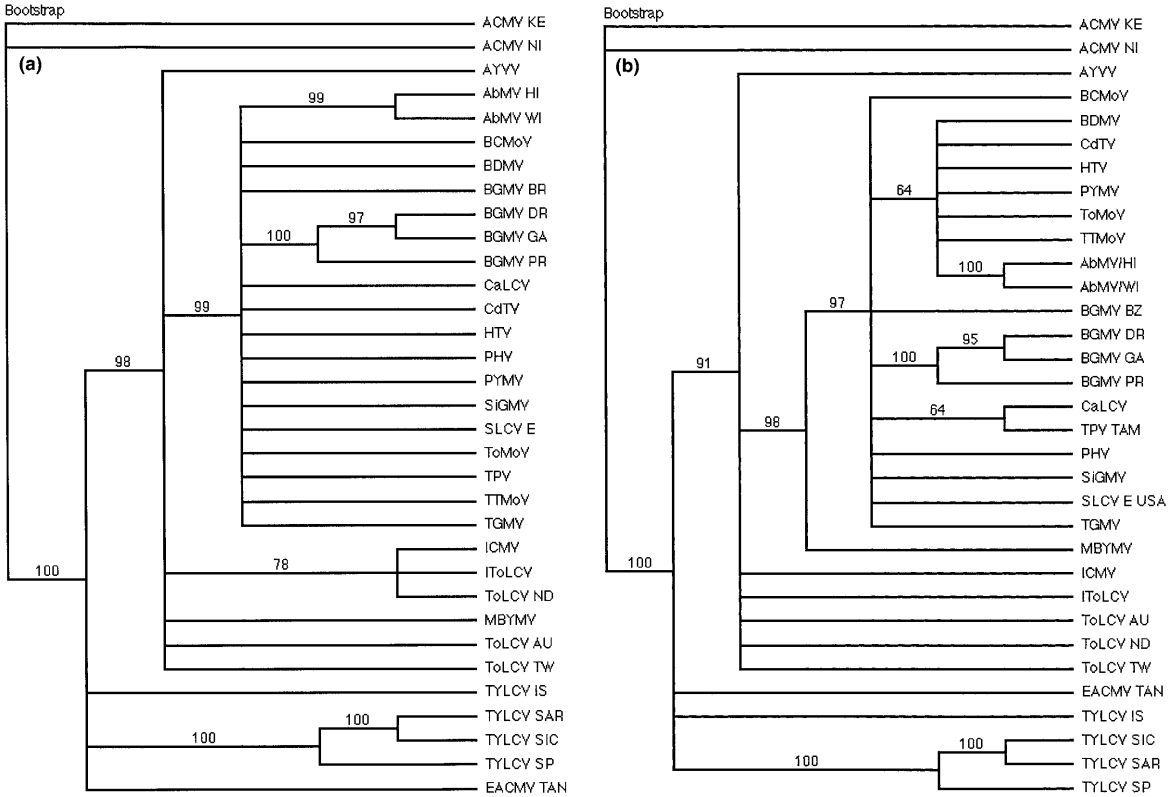

Fig. 1. Cladogram illustraging relationships between representative begomoviruses based on (a) *core CP* and (b) *CP* gene sequences. Trees were reconstructed by aligning sequences with Clustal followed by maximum parsimony analysis using PAUP 3.1.1. A most parsimonious tree was produced from 200 bootstrap iterations using the heuristic search, TBA, and random sequence addition options. Nodes are collapsed for branching topologies taht did not occur at least 60% of the time

field isolates, the *core CP* is demonstrated to be readily accessible and competent for achieving preliminary virus identification that should be followed by evaluation of other key sequences. *Core CP* comparisons illustrate the feasibility of comparative analysis between an 'unknown' and reference sequences in predicting placement of the unknown, in the context of well-characterized, reference begomoviral sequences. Moreover, pairwise distance estimates or extent of sequence divergence between isolates/viruses facilitate a quantitatively based prediction of virus identity ($>90\%$), or identification of a closest begomovirus relative ($<90\%$) (data not shown).

Based on *core CP* sequence alignment and the resultant cladogram, reference viruses are readily separated into groups based first on insect vector in that the two Cicadellidae-transmitted virus clusters are separated (data not shown) from viruses of the large *B. tabaci*-transmitted group (Fig. 1). Within the whitefly-transmitted clade, there is a clear separation between viruses of Eastern and Western Hemisphere lineages (Fig. 1). It is notable that even with the subsequent addition of the many sequences obtained from field isolates (Fig. 2), there was no compelling evidence for additional phylogenetic clusters. Hence, based on the *core CP* sequence analysis, it is possible to predict when viruses are (recently) indigenous to the Old or New World, and, for some, to discern an affiliation with a specific region within that hemisphere, a feature that is particularly notable for the Eastern Hemisphere viruses and less clear for viruses from The Americas and Caribbean. This approach permits tentative assignment of a new, previously unreported virus to a particular virus sub-cluster, indicating its nearest relative and making tentative

identification possible based on a closest match. This tool further permits the rapid recognition of most exotic viruses in a timely manner.

The availability of universal PCR primers for amplification, the ease and availability of automated sequencing and the establishment of an interactive web-based database makes it possible to identity rapidly a begomovirus sample via the internet. A large *CP-core* sequence database exists for many isolates, including laboratory-authenticated and field isolates representing a broad geographical range and diverse host species. Begomovirus *core CP* sequences are available through GenBank and at the GEMINIDETECTive website, both sites permitting users to identify an unknown begomovirus utilizing BLAST and taking only a few minutes online to complete.

3.2. Comparison of predictions by core CP and CP nucleotide sequences

The *CP* gene is the most highly conserved gene within the *Geminiviridae* (Harrison, 1985; Padidam et al., 1995; Rybicki, 1994) and has been accepted as a valid sequence marker for establishing the provisional identity of begomoviruses when a full length viral genome sequence is not available (Mayo and Pringle, 1998). In a previous comparative analysis of begomovirus sequences, nearly congruent trees were predicted by the full-length monopartite/A component viral sequences, complete *CP* gene sequence, 5′-proximal 200 *CP* sequences described by Padidam et al., (1995), and *core CP* sequences (Brown et al., submitted for publication).

Using *core CP* or *CP* sequences, trees were generally congruent in predicting relationships among begomoviruses (Fig. 1a, b , respectively), but several anomalies were observed. The *core CP* tree (Fig. 1b) separated Eastern and Western Hemisphere viruses into distinct clusters containing (1) all AbMV isolates, (2) all Caribbean bean-infecting isolates of BGMV, (3) a large polytomy containing all other New World viruses, (4) tomato-infecting viral species and strains from the Mediterranean region, (5) three of four species from India, and (6) a large polytomy containing all other Old World viruses. In contrast, the *CP* gene tree predicted several relationships among New World viruses not revealed in the *core CP* polytomy: (1) a group containing AbMV plus BDMV, CdTV, HTV, PYMV, ToMoV and TToMoV, and (2) a cluster containing CaLCV and TPV. Also, the *CP* tree separated the three viruses from India as distinct species within a large Old World polytomy. Based on pairwise comparisons of *CP* nt sequences, the two most divergent New World viruses were BGMV-PR and AbMV-WI (28%), while the two most divergent Old World taxa were TYLCV-Sar and TYLCV-Th (29.5%). All other begomoviruses diverged to a lesser extent, and shared sequence identities ranging between 73 and 89% for recognized species, while strains shared 95–99% identity. New World viruses were more closely related to one another than they were to any Old World virus, while Old World viruses were likewise most closely related to their own geographic counterparts.

Despite several minor inconsistencies between *core CP* and *CP* trees, between taxa divergence values were generally comparable for the two markers, confirming that *core CP* sequences enable predictions about begomovirus relationships at the level of strain versus species that are consistent with predictions based on the *CP* sequence (providing recombination has not occurred within 'untargeted' regions of the gene (3′ or 5′ end). The highly variable nature of the 5′ end of the *CP* suggests that the absence of this hypervariable region in the *core CP* marker sequence may alone be responsible for the inability of the *core CP* to resolve an AbMV-*CP* cluster. This observation may also reflect 'noise' in the *core CP* owing to the larger proportion of variable to conserved nt which is slightly higher than for the *CP*. Further, the influence on divergence estimates of the large number of highly variable nt positions at the 5′ end of the *CP* may be offset by the highly conserved 3′ end, the latter and former segments being included and absent, respectively, in the *core CP*. Nevertheless, a greater proportion of variable sites may actually reveal unique insights into evolutionary patterns, particularly for less divergent taxa (including strains), owing to the greater proportion of sites that can vary in the *core CP* compared to the entire *CP* sequence.

Collective observations (Padidam et al., 1995; Rybicki, 1994) based on sequence divergences estimates suggested that begomovirus species in the Old World were generally less closely related than are species from the New World. However, as more New World viruses are discovered, views regarding begomovirus genealogy and evolutionary history may become better understood. Certainly, it is far more difficult to ascribe a specific geographic origin to most New World taxa. The majority of New World begomoviruses are more readily separated by genotypic lineage than by discrete geographic region as are old world viruses. Extant patterns of begomovirus distribution are none the less suggestive of an early and complete separation of Old and New World viruses. Support for the hypothesis that begomoviruses have an Old World origin can be drawn from the prediction that their vector (*B. tabaci*) originated from the Indian Sub-continent or Africa (Gill, 1992b; R. Gill, personal communication). Thus, genotypes that originated from the Old World are the most likely predecessors of extant New World begomoviruses. Consequently, predecessors of the New World group became established during one or more 'founder events' involving reasonably divergent genotypes, based on the contemporary range of divergence between New World subgroup members. Perhaps the somewhat greater extent of divergence between certain Old World groups of begomoviruses suggest a longer history of isolation, possibly with a basis in a more absolute separation between geographically- or host-isolated Old World taxa, than has been permitted amongst the New World members. That these separations occurred some time ago is substantiated thus far because there are no obvious 'bridging' taxa linking unequivocally the most disparate Old World groups, or New and Old World clades.

3.3. Comparative analysis of CP ORF sequences of field isolates

To extend and validate the utility of *CP* sequences as a marker for predicting viral relationships 80 *CP* sequences were compared, including 42 from field isolates and 40 reference sequences (Fig. 2). From this analysis, it can be demonstrated that several previously undescribed begomoviruses were discovered and that numerous previously known viruses were documented in either new or expected locations. Estimated divergence ranged from 1.9% to 33.5% for field and reference virus sequences: New World viruses/strains diverged by $\sim$10–22% and 18–30% respectively between those from the Old World (data not shown).

Clusters of New World viruses were grouped at times with a basis in genus or plant family of the host from which they were collected. Several examples include isolates from cotton/*Sida* (Malvaceae), pepper, tomato (Solanaceae) and cassava (Euphorbiaceae). In contrast, the majority of isolates did not share a common host affiliation, but rather a geographic affiliation (Fig. 2). Most isolates/viruses grouped according to one or several geographic locations in either the New or Old World. A single example was seen in which a virus isolate, identified here as TYLCV-Is, originates from the Middle East, but was collected from a New World site. TYLCV-Is is known to have been introduced to the Caribbean region in 1991–1992 (J.K. Brown and J. Bird, unpublished report; Polston et al., 1994).

CP gene trees for the New World viruses indicated an overlap in distribution of species and strains from one to three main regions: (1) the southern 'sunbelt' states of the US and northern/southern Mexico, (2) the Caribbean region and Central America, (3) the Caribbean region (legume-infecting), (4) Caribbean/Central American regions and South America, (5) southern Mexico and Central America, and (6) South America (Fig. 2). Results suggest that important geographic barriers to the dissemination in the New World were rare and/or have not generally influenced the distribution of begomovirus lineages either naturally over time. In contrast, the Eastern Hemisphere begomoviruses were grouped mainly by affiliation with single or adjacent geographic regions. The *CP* nt tree reveals groups of Old World begomoviruses centred in the (1) Indian Sub-continent, (2) region comprising the boundaries of the Mediterranean Sea: Iberian

Peninsula, Middle East and Northern Africa, (3) sub-Saharan and continental Africa with some evidence of host separation (tomato vs. cassava), and (4) The Far East/Pacific Rim and Australia. These latter relationships are consistent with an argument for more numerous and isolated lineages of Old World viruses, compared to New World viruses. Additional studies may clarify some of these relationships and could reveal patterns of recombination or reassortment, as well as

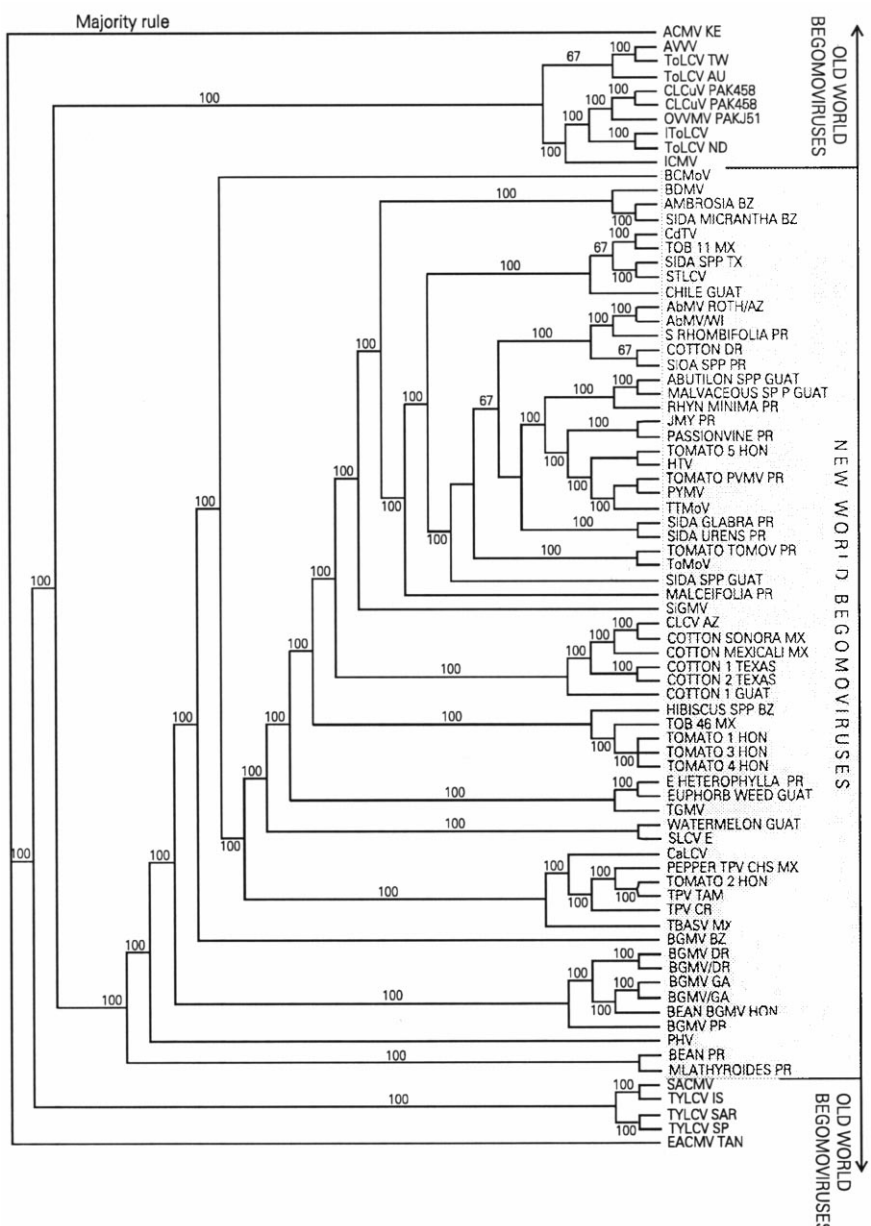

Fig. 2. Tree showing the most parsimonious arrangement for 80 begomoviruses based on the nt sequence of the *CP* gene (200 replicates), employing the $>60\%$ majority rule and Heuristic Search Option in PAUP 3.1.1.

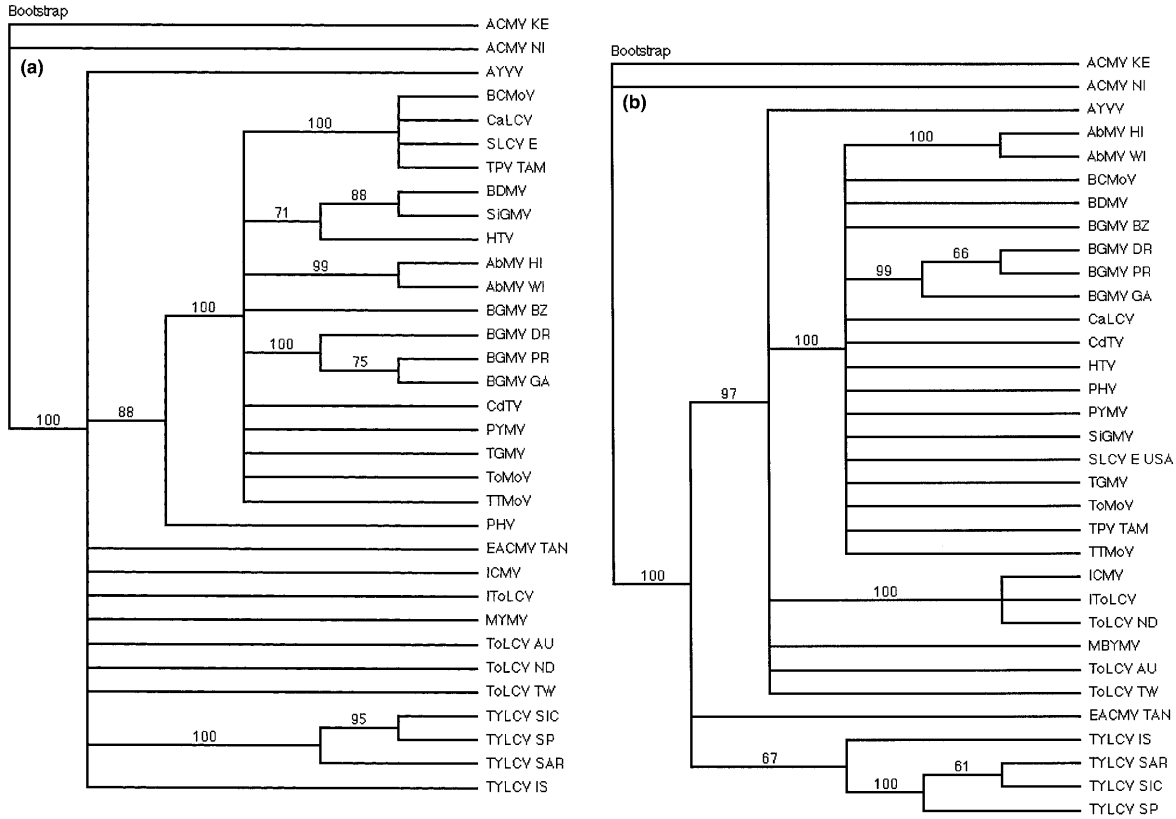

Fig. 3. Cladogram illustrating relationships between the amino acids of the (a) replication associated protein (L1/C1 ORF) and (b) coat protein (V1 ORF) of representative begomoviruses. Trees were reconstructed by aligining amino acid sequences with Clustal. A most parsimonious tree was produced from 200 bootstrap iterations using the heuristic search, TBA, and random sequence addition options in PAUP 3.1.1. Nodes are collapsed for branching topologies taht did not occur at least 60% of the time.

lineages that may have arisen in isolation owing instead to host plant and/or whitefly vector restrictions.

3.4. Comparison of AA sequences of REP and CP ORFs

The utility of the N-terminal 200 AA of the REP (associated protein) polypeptide (Fig. 3a) has been examined as a prospective informative sequence and the results were compared to those for the *CP* AA sequence (Fig. 3b) the nt sequence of *CP* having been accepted by the ICTV for establishing provisional virus identification. This region of REP was selected because it is characterized by the presence of highly conserved AA positions (Fig. 4) that are known to be virus-specific (Hanley-Bowdoin et al., 1999) and perhaps lineage-associated in some instances.

The REP polypeptide tree revealed relationships similar to those inferred by the *CP* tree, with several exceptions (Fig. 3a, b). Most taxa were placed on REP and *CP* AA trees, as expected, by geographic origin in the Old or New World. REP and *CP* tree both predicted a New World subgroup that contained two AbMV isolates (Hi, WI), and the three Caribbean bean-infecting isolates (DR, Ga, PR). while the REP comparisons uniquely REP predicted a SLCV cluster containing BcMoV, CaLCV, TPV-Tam, SLCV-E, BDMV and SiGMV. Analysis of complete genomic or *CP* nt sequences typically place the latter two viruses in a different group (Fig. 1b, Fig. 2). Among Old World viruses, REP predicted

a polytomy containing all taxa except those from the Iberian Peninsula, which are typically grouped with TYLCV-Is based on *CP* or complete genomic sequence. Further, the *CP* AA tree resolved a unique group containing three of the four viruses from India (Fig. 3b), the same pattern seen for the *CP* tree (Fig. 1b), whereas the REP AA tree did not distinguish this cluster from other taxa within the Old World polytomy. Sequence divergence estimates for AA REP, AA *CP*, and nt *CP* based on pairwise comparisons are generally consistent with respect to the range of inter-taxa divergences, but AA REP is overall more divergent than AA *CP* (data not shown), as noted previously (Harrison, 1985).

```
                                        Helix I  Loop   Helix II
            1             I      ***********XXXXX^^^^^^^^^^^    II 61
AbMV Hi     MPPPEKFRVQAKNYFLTYPQYSLAKEVALSQLQNLETPVNKKFIEICRELHENGEPHLHV..L
AbMV WI     MPPPKKFRVQAKNYFLTYPQCSLTKDEALSQLQNLETPVNKKFIKICRELHENGEPHLHV..L
ACMV Ke     MRTPR.FRIQAKNVFVTYPKCSIPKEHLLSFIQTLSLQSNPKFIKICRELHQNGEPHLHA..L
ACMV Ni     MRTPR.FRVQAKNVFLTYPNCSIPKEHLLSFIQTLSLPSNPKFIKICRELHQNGEPHLHA..L
AYVV        MAPPKRFQINAKNYFLTYPQCSLTKEEALSQLQNLNTPTNKKYIKICRELHEDGSPHLHV..L
BCMoV       MPRNPSSFRITARNIFLTYPQCDINKDEALRLLQELPWSVVKPTYIRVAREEHSDGFPHLHCL
BDMV        MPPPKKFRVQSRNYFLTYPQCSLTKEEALSQIQNLKTPVNKKFIKICRELHEDGEPHLHV..L
BGMV Dr     MPSPHRFRVQSKNYFLTYPHCSIPKEEALSQLQKIHTATNKKFIKVCEERHENGEPHLHA..L
BGMV Ga     MPPPQRFRVQSKNYFLTYPRCPIPKEEVLSQLQKIHTATNKKFIKVCEERHENGEPHLHA..L
BGMV PR     MPPPQRFRVQSKNYFLTYPRCTIPKEEALSQLQKIHTTTNKKFIKVCEERHDNGEPHLHA..L
BGMV-Bz     MPPPKRFKINAKNYFLTYPQCSITKESAIEQLQNLQTPVNKKYIRICREIHENGEPHLHA..L
CaLCV       MPRNPKSFRLAARNIFLTHLRCDIPKDEALQMLQTLSWSVVKPTYIRVAREEHSDGSPFTHCL
CdTV        MPSVKRFKVSAKNYFLTYPQCSLSKEEALSQLQTLKTPVNKKYIKICRELHENGEPHLHV..L
EACMV Tan   MPRAGRFQINAKNYFITYPRCSLTKEEALSQLKALSYPTNIKFVRVCRELHHDGVPHLHV..L
HTV         MPPPKKFRVNSKNYFLTYPQCSITKEEALSQIKNLNTPVNKKFIKICRELHENGEPHLHV..L
ICMV        MSPPKRFQINAKNYFLTYPRCSLTKEEALSQIRNFQTPTNPKFIKICRELHENGEPHLHV..L
ITOLCV      MPRPGRFNINAKNYFLTYPKCSLSKERHLSQLQTVKTPTSKLFIRVCRELHVDGEPHLHV..L
MYMV        MPRLGRFAINAKNYFLTYPRCPLRKEDALEELLALSTPVNKKFIRVCRELHEDGEPHLHV..L
PHV         MPLPKRFRLNAKNYFLTYPQCSISKEERLAQLQNLSTPVNKKYIKICKESHEDGQPHLHV..L
PYMV        MPRKGSFSIKAKNYFLTYPQCSISKEEALSQLQNLTIPVNKKFIKICRELHENGEPHLHV..L
SiGMV       MPPPKKFRVQSKNHFLTYPQCSLTKEEALSQLQNLATPVNKKFIKICRELHEDGQPHLHV..L
SLCV E      MPRNPNSFRLTARNIFLTYPRCDVPKEEVLEMLLHLSWSVVKPNYVRVAREEHSDGSPHLHCL
TGMV        MPSHPKRFQINAKNYFLTYPQCSLSKEESLSQLQALNTPINKKFIKICRELHEDGQPHLHV.L
ToLCV Au    MTRPKSFRINAKNYFLTYPKCSLTKEEALSQLNNLE.TPTSKKYIKVCRELHENGEPHLHV.L
ToLCV NDe   MAPPRRFRIDAKNYFLTYPKCSLTKEEALSQLQTLE.TPTAKKFIKICRELHDDGSPHIHV.L
ToLCV Tw    MHPLNKFRINAKNYFLTYPHCSLTKEEALSQIQALQ.TPTNKLFIRVCRELHEDGAPHLHI.L
ToMoV       MPPPKKFRVQSKNYFLTYPQCSLSKEEALSQLQNLN.TPVNKKFIKICRELHENGEPHLHV.L
TPV Tam     MPLPPKSFRLQCKNIFLTYPQCDIPKDEALEMLRNLKWAVVKPIYLRVSREEHSDGFPHLHCL
TTMoV       MPRKGSFSVKAKNYFLTYPQCSLSKEEALSQIQNLN.TPVDKKFIKICRELHENGEPHLHV.L
TYLCV Is    MPRL..FKIYAKNYFLTYPNCSLSKEEALSQLKKLE.TPTNKKYIKVCKELHENGEPHLHV.L
TYLCV Sar   MPRSGRFSIKAKNYFLTYPKCDLTKENALSQITNLQ.TPTNKLFIKICRELHENGEPHLHI.L
TYLCV Sic   MAQPKRFQINAKHYFLTFPKCCLTKEEALEQLLKLQ.TPTNKKYIKICRELHEDGQPHLHI.L
TYLCV Sp    MAQPKRFQINAKHYFLTFPKCSLSKEEALEQLLQLQ.TPTNKKYIKICRELHEDGQPHLHM.L

            62                                      III                 12
AbMV Hi     IQFEGKFQCTNYRFFDLVSPTRAAHFHPNIQGAKSSSDVKSYIDKDGDTAEWGEFQIDGRSTRG
AbMV WI     IQFEGKYQCTNNRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDTAEWGEFQIDGRSARG
ACMV Ke     IQFEGKITITNNRLFDCVHPSCSTSFHPNIQGAKSSSDVKSYLDKDGDTVEWGQFQIDGRSARG
ACMV Ni     IQFEGKITITNNRLFDCVHPSCSTNFHPNIQGAKSSSDVKSYLDKDGDTVEWGQFQIDGRSARG
AYVV        MQFEGKYKCQNNRFFDLVSPSRSAHFHPNIQGAKSSSDVKSYIDKDGDTLEWGEFQIDGRSARG
BCMoV       IQLSGKSNIKDVRFFDLTHPRRSANFHPNVQAAKDTNAVKNYITKEGDYCESGEYKVSG.S...
BDMV        IQFEGKYQCTNNRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDTVEWGVFQIDGRSARG
BGMV DR     IQFEGKFVCTNKRLFDLVSSTRSAHFHPNIQGAKSSSDVKAYIDKDGDTIEWGQFQVDGRSARG
BGMV Ga     IQFEGKFVCTNKRLFDLVSSTRSAPFHPNIQGAKSSSDVKAYIDKDGVTIEWGQFQVDGRSARG
BGMV PR     IQFEGKFICTNKRLFDLVSTTRSAHFHPNIQGAKSSSDVKEYIDKDGVTIEWGQFQVDGRSARG
```

Fig. 4. Alignment of the 5′ ∼189 AA of the Rep associated protein (AL1/C1 ORF) for begomoviruses representing the Eastern and Western Hemispheres. Shown here are the positions of the three N-terminal motifs found in the replication protein, I (DNA cleavage?), II (metal-binding site), and III (contains conserved tyrosine residue required for DNA cleavage and ligation), the former conserved in the Rep protein of all begomoviruses and the latter in begomovirues and the initiator proteins of other rolling circle systems of replication. Functions assigned to the N-terminal half of the polypeptide include the site of AL1/C1 oligomerization (two predicted alpha-helices sets, I, II, and III, IV), virus-specific binding to *cis* element (direct repeat) in the intergenic/CR, and DNA cleavage and ligation. The ATP/GTP-binding site is located within a fourth recognized motif located in the C-terminal portion of the polpeptide, and was not included in this alignment.

```
BGMV-Bz     IQFEGKFQCTNCRVFDLKHPTTSSVSHANIQSAKSSSDVKSYIEKDGDYIEWGHFQVDGRSARG
CaLCV       IQLSGKSNIKDARFFDITHPRRSANFHPNIQAAKDTNAVKNYITKDGDYCESGQYKVSG....G
CdTV        IQFEGKYQCTNNRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDTVEWGEFQIDGRSARG
EACMV Tan   IQFEGKFQCTNPRFFDLISPSRSTHFHPNIQGAKSSSDVKAYIEKGGEFLDDGIFQVDARSARG
HTV         LQFEAKYQCTNSRFFDLVSPSRSAHFHPNIQGAKSSSDVKAYIDKDGDATEWGEFQIDGRSARG
ICMV        IQFEGKYKCQNQRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDTWRWGTFQIDGRSARG
ITOLCV      IQFEGKFQCTNNRFFDLVSPTRSTHFHPNIQKAKSSSDVKAYVEKDGDFIDFGVFQIDGRSARG
MYMV        LQFEGKFQTKNERFFDLVSSTRSAHYHPNVQAAKSASDVKSYMDKDGDVVDHGSFQVDGRSARG
PHV         IQFEGKYQCTNNRFFDLVSSTRSAHFHPNIQGAKSSSDVKTYIDKDGDTVEWGEFQIDGRSARG
PYMV        IQFEGKYNCTNNRLFDLVSPTRSTHFHPNIQGAKSSSDVKSYVEKDGDTIEWGLFQIDGRSARG
SiGMV       IQFEGKYQCTNQRFFDLVSPTRSVHFHPNIQGAKSSSDVKSYIDKDGDTVEWWEFQIDGRSARG
SLCV E      IQLSGKSNIKDAGFFDLTHPRRSARFHPNIQAAKDTNAVKNYITKEGDYCESGQYKVSG....G
TGMV        IQFEGKYCCQNQRFFDLVSPTRSAHFHPNIQRAKSSSDVKTYIDKDGDTLVWGEFQVDGRSARG
ToLCV Au    IQFEGKFQCKNQRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYLEKDGDTLEWGEFQIDGRSARG
ToLCV NDe   IQFEGKFQCKNNRFFDLVSPSRSAHFHPNIQGAKSASDVKAYIDKDGDVLEWGVFQIDGRSARG
ToLCV Tw    IQFEGKFQCKNNRFFDLTSPTRSAHFHPNIQGAKSSTDVKTYMEKDGDVLDHGIFQVDGRSARG
ToMoV       VQFEGKYQCTNNRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDTIEWGDFQIDGRSARG
TPV Tam     IQLTGKCNIKDARFFDITHPRRSAQFHPNVQASKDANAVKNYITKDGDYCESGQYKVSG....G
TTMoV       VQFEGKYNCTNNRFFDVVSPTRSAHFHPNIQGAKSSSDVKTYITKDGDTLEWGEFQIDGRSARG
TYLCV Is    IQFEGKYQCKNQRFFDLVSPNRSAHFHPNIQAAKSSTDVKTYVEKDGNFIDFGVSQIDGRSARG
TYLCV Sar   IQFEGKYNCTNQRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDVLEWGTFQIDGRSARG
TYLCV Sic   MQFEGKFNCKNNRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDVLEWGTFQIDGRSARG
TYLCV Sp    IQFEGKFNCKNNRFFDLVSPTRSAHFHPNIQGAKSSSDVKSYIDKDGDVLEWGTFQIDGRSARG

                    Helices 3&4
            126  >>>>>>>>>>>>>>>>>>>>>>>>                                  189
AbMV Hi     GQQTANDSYAKALNAIDVQSALNILKEEQPKDYVLQNHNIRSNLERIFAKAPEPWVPPFPLSSF
AbMV WI     GQQTANDSYAKALNAGDVQSALNILKEEQPKDYVLQNHNIRSNLERIFAKAPEPWVAGFPSPLS
ACMV Ke     GQQSANDAYAKALNSGSKSEALNVIRELVPKDFVLQFHNLNSNLDRIFQEPPAPYVSPFPCSSF
ACMV Ni     GQQSANDAYAKALNSGSKSEALNVIRELVPKDFVLQFHNLNSNLDRIFQEPPAPYVSPFPCSSF
AYVV        GQQTANDAYAQALNSGSKSEALNVIKELAPKDYVLQFHNLNANLDRIFAPPLEVFVCPFLSSSF
BCMoV       TKSNKDDVYHNAVNAGSIKEALDIIRAGDPKTFIVNYHNIKANIERLFKKAPEPWVPPFQLSSF
BDMV        GQQSANDTYAKALNAGSAEQALRIIKEEQPQHFFLQHHNLVANATSIFKKAPEPWVPPFPLSSS
BGMV DR     GQQTANDSYAKALNADSIESALTILKEEQPKDYVLQNHNIRSNLERIFFKVPEPWVPPFPLSSF
BGMV Ga     GQQSANDSYAKALNADSIESALTILKEEQPKDYVLQHHNIRSNLERIFVKVPEPWVPPFPLSSF
BGMV PR     GQQSANDSYAKALNADSIESALTILKEEQPKDYVLQNHNIRSNLERIFFKVPEPWVPPFPLSSF
BGMV-Bz     GQQTANDAASEALNASSKEEAMQIIKEKLPEKFLFQYHNLSSNLDRIFTKAPDPWSPPYHLSSF
CaLCV       TKANKDDVYHNAVNAGCVEERLAIIRAGVPKTFIVSYHNVRANIERLFTKAPEPWAPPFQLSSF
CdTV        GQQSANDSYAKALNAGSVQAAMTVLKEEQPKDFVLHNHNIRSNLERIFAKAPERWVPPFQLSSF
EACMV Tan   EGQHLAQVYAEALNASSKSEALQIIKEKDPKSFFLQFHNISANADRIFQAPPQTYVSPFLSSSF
HTV         GCQSANDTYAKALNASCAEEALQIIKEQQPQHFFLHHHNLVANAQRIFHKPPEPWVPPFHLSSF
ICMV        GQQSANDAYAAALNSGSKSEALKILRELAPRDYLRDFHHISSNLDRIFTKPPPPYENPFPLSSF
ITOLCV      GQQSANDAYAEAINSGSKAAALDILREKAPKDFVLQFHNLNANLDRIFTPSAEVYVSPFLCSSF
MYMV        GKQSANDAYAEALNSGSKLQALNILREKAPKDYILQFHNLNCNLSRIFSDEVPLYVSPYSLSAF
PHV         GQQSANDTYAKALNSASAEEALQIIKEEQPQHFFLQFHNIVSNANRIFQTPPEPWVPPFQQASF
PYMV        GQQTVNDAAAEALNSGTKEAAMKIIKEKLPEKFLFQYHNLSCNLDRIFMKAPETWAPPFPLSSF
SiGMV       GQQSANDSYAKALNASSAEEALRIIKEEQPQHFFLQHHNLVANATRIFQKAPSPWAPPFPLSSF
SLCV E      SKSNKDDVYHNAVNAGSAGEALDIIKAGDPKTFIVNYHNLLANVERLFQKPPEPWVPPFELSSF
TGMV        GCQTSNDAAAEALNASSKEEALQIIREKIPEKYLFQFHNLNINLDRIFDKTPEPWLPPFHVSSF
ToLCV Au    GQQSANDAYAQALNTGSKSEALNVLRELAPKDYVLQFHNLNSNLDRIFTPPLEVYVSPFLSSSF
ToLCV NDe   GQQTATDAYAKAINTGNKEDALKVLKELAPKDYVLQFHNLISNLDRIFQPRSEVYVSPFSISSF
ToLCV Tw    GCQSANDAYAEAINSGSKAEALNILREKAPKDFVLQFHNLNSNLDRIFTPPLEVYISPFNSSSF
ToMoV       GQQSANDSYAKALNAGSVQSALAVLREEQPKDFVLQNHNIRSNLERIFAKAPEPWVPPFQVSSF
TPV Tam     SKANKDDVYHNAVNAASAGEALDIIKAGDPKTFIVSYHNIKANIERLFTTPPMPWTPPYPLSSF
TTMoV       GQQSANDSYAKALNAGNVQSAMAVLKEEQPKDFVLHNHNIRSNLERIFAKAPEPWVPPFPLSSF
TYLCV Is    GQQSANDAYAEALNSGSKSEALNILKEKAPKDYILQFHNLSSNLDRIFSPPLEVYVSPFLSSSF
TYLCV Sar   GQQTANDAYAKAINAGSKSQALDVIKELAPRDYVLHFHNINSNLDKVFQVPPAPYVSPFLSSSF
TYLCV Sic   GQQTANDAYAKAINARSKSEALDVIKQLAPRDYVLHFHNISSNLDKVFQVPPAPYVSPFLSSSF
TYLCV Sp    GQQTANDAYAKAINAGRKSEALDVIKELAPRDYILHFHNINSNLNMVFQVPPAPYVSPFLSSSF
```

Fig. 4. (*Continued*)

The significance of these findings in the context of making predictions concerning the capacity for cross-replication between distinct viral species or strains will require empirical experimentation employing full-length infectious clones. This is because the particular sequences and structural considerations involved in the making of an ideal, viable virus have not yet been elucidated satisfactorily. Mean-

while, perhaps such simplistic predictions will enable the selection of a strategic suite of viral genotypes that when employed in reassortment experiments will assist in defining the functional genomics of begomoviral sequences involved in the sustainability of viable reassorted and/or recombined combinations of non-cognate begomoviral components.

3.5. Survey of IR/CR elements that bind REP-associated protein

Here we identified and compared the composition of the directly or inverted repeated sequences or 'iterons' located in the IR/CR of begomoviruses (Arguello-Astorga et al., 1994). These are sequences shown recently to constitute elements involved in virus-specific binding of REP during initiation of viral replication (Hanley-Bowdoin et al., 1999). Directly repeated elements were identified for begomoviruses for which sequences are available in GenBank, or from begomovirus isolates being investigated in our laboratory (Table 1).

A New World begomovirus iteron consists typically of 5 nt (G- or GT/A-rich) at the 5′ end of the element. This segment is separated from the 3′ direct repeat (which is not always identical to the 5′ 'repeat') by a 2–4 nt 'spacer' composed of A, T, and/or C residues (Table 3). Iterated sequences of Old World begomovirus elements are more difficult to identify, are often more cryptic in location and typically also include an inverted repeat (data not shown). In many instances, the 5′ and 3′ segments of the REP-binding element were not identical and segments within an element varied by 2–4 nt. In one unusual example, the 5′ and 3′ segments of the iterative element for the bean-restricted BGMV-Caribbean group are different from each other, and instead of the typical 5 nt length, segments were only 4 nt long and were separated by a dinucleotide (AC) instead of a tri-nucleotide spacer. Moreover, ACMV-Ke and ACMV-Ni contained a 9 nt spacer sequence and distinct 5′ and 3′ direct repeats. In general, elements were virus-specific, although several viruses shared similar or identical REP-binding sites (Table 3).

As to be expected from an alignment of short sequences, inferred viral genealogies (data not shown) were incongruent with all other nt or AA cladograms, nor were groups associated with geographic distribution or host. The notion of employing 'direct repeat families' in predicting lineage associations is feasible when patterns are apparent, as in members of the SLCV group, for example. However, when sequence elements have diverged beyond recognition, or are present in genotypes with no apparent recent contact or predicted evolutionary histories, this approach is not particularly insightful. Initially, we were surprised at the extent of variation seen in the second segment of many of these elements, though in retrospect, this was unwarranted, owing to the recent discovery that the first elemental segment is essential for REP binding with TGMV (Hanley-Bowdoin et al., 1999).

Several striking examples of 'solanaceous-associated relatives' that share sequence identity in the leftward elemental segment are: GGGGG for TYLCV-Sar, TToMoV (Cuba), and ToLCV- (India); GGTAG for TGMV (Brazil), TYLCV-Sic, and TYLCV-Sp; GGTGT for tobacco apical stunt virus (Mexico) (Paximadis et al., 1999), PHV, ToLCV-Tw (Taiwan), TYLCV-Is, ToLCV-Au, and ToLCV-NDe (albeit, several non-tomato-infecting viruses also share this elemental sequence, including the SLCV lineage from the Americas and Caribbean).

The best example of a lineage for which the direct repeat sequence is shared in common is seen for the SLCV group (GGTGT) whose members also share to a great extent geographic origin and host range. This cluster includes SLCV-E, SLCV-R, BCaMV, CaLCV, TPV, and the recently described *Cucurbit leaf curl virus* (CuLCV) (Brown et al., 2000b). Interestingly, the only virus in the lineage that does not infect cucurbits and bean is TPV, suggesting a cryptic, intriguing link. Indeed, the ability to participate in cross-replication and movement has been documented for certain combinations of SLCV-E and CaLCV (Hill et al., 1998).

A particularly conflicting example may be taken from reassortant experiments with ToMoV (Florida, US) and BDMV (Colombia). These two

tomato-infecting viruses which have different efficiencies of replication in tomato and non-identical REP-binding element sequences (Table 3) were shown to be capable of forming reassortants, albeit in one direction. Despite extremely low rates of infection, these two viruses could cross-replicate in tomato (Hou and Gilbertson, 1996). Clearly, more than the specific nt sequence of the REP-binding element governs replication specificity and the viability of begomoviruses as successful pathogens.

Nevertheless, that direct or inverted repeats appear relatively conserved within a species and at times a lineage, selection apparently favours the retention of at least certain of the rightward elemental sequence, perhaps owing more to structural considerations. Though speculative, it is possible that begomoviruses can inherit the leftward segment and spacer and as a modular unit during recombination. Whether it is likely that 'repeated' portions of an element frequently become isolated during recombination is not known. However, it is conceivable that a 'new virus' might arise which contains segments from two different elements, given a compatible REP. Likewise, when Old World lineages gave rise to viruses now extant in the New World, it is possible that REP-binding elements were strongly conserved, even if other regions of the genome have since diverged. Interestingly, several Old and New World tomato-infecting begomoviruses share the same direct repeat (Table 3). It is also feasible that directly repeated elements arose through interactions with foreign genomes. Elements could have been 'inherited' together with the nearby consensus nonanucleotide that is also present in other 'life forms' that utilize ssDNA rolling circle replication (Ilyina and Koonin, 1992; Stanley et al., 1995).

3.6. Genic variation within vector populations

Accurate identification of insects that are pests and/or virus vectors is a prerequisite for their effective management to reduce crop damage. An understanding of relevant biological aspects of many insects, irrespective of their pest status, is often thwarted by unrecognized sibling species that exhibit biological differences. This is particularly important for species that are rich in biological variants that lack distinguishing morphological characteristics. A tropical insect pest and timely plant virus vector for which this holds true is *B. tabaci*, which is a species synonymized from several genera and 18 species (Russell, 1957).

Although molecular markers for separating *B. tabaci* variants within the complex have become available only recently (De Barro et al., 2000; Frohlich et al., 1999), interesting biogeographic relationships have been revealed in the preview that are consistent with the origin of *B. tabaci* from the Old World, either in India or Africa, the two continents that harbour diverse natural enemies (Mound and Halsey, 1978; Kirk et al., 2000).

Here, we provide a purview into the relationships between a collection of *B. tabaci* from locations, worldwide (Table 2) based on phylogenetic analysis of the mt COI and note that there are striking congruencies, though in retrospect, perhaps not surprising, between the begomovirus *CP* tree and the whitefly vector mt COI trees. Analysis of mt COI sequences revealed five main groups of *B. tabaci* with a basis primarily on geographic origin (Fig. 5). Relationships revealed by the cladogram group all collections from the New World as one large lineage exhibiting minor degrees of divergence (data not shown), while in contrast, four or five Old World groups were delimited: (1) Southeast Asia: Thailand, Malaysia, and Pakistan (2) two subgroups mainly from areas around the Mediterranean Sea: subgroup A, Spain; Israel, and Northern Africa; and Zimbabwe; and subgroup B, the 'B biotype' (found in many sites worldwide) and either its closest relatives from Israel and adjacent locales (B type prior to bottleneck?), (3) India, Pakistan, Nepal (distinct from group 1), and (4) sub-Saharan Africa: Cameroon, Mozambique, Uganda, and Zambia (Fig. 5).

The large group from the Mediterranean region contains two subgroups, one that harbours only the 'B'-type, now having a widespread international distribution and a related but distinct group of 'non-B' variants from Spain and Sudan. Placement of the invasive 'B-type' within this larger

group is consistent with evidence that *B. tabaci* type B was disseminated throughout the world on ornamental crops most recently originating from Israel, yet the precise origin of the B-type is still uncertain. These data suggest the B-type's closest relatives reside amongst the large Mediterranean region group (Frohlich et al., 1999). Interestingly, two Old World groups containing lineages from the Indian Sub-continent: India–Nepal–Pakistan (PC 92) and Pakistan (PC 91, PC 95)-SE Asian groups, a finding also substantiated by rather substantial divergence (*c*.11%) between these two groups (data not shown). Given that the Indian Sub-continent may be the centre of origin of *B. tabaci*, much variability would be expected among collections from this region. Likewise, Africa is also proposed as a centre of diversity for *B. tabaci*. Although African groups have only been explored peripherally (Rey et al., 2000), a wide range of variability has been observed among the populations examined (data not shown). Clearly, additional sampling and more comprehensive studies will provide insights into the extent of variation and range of vector variants in both postulated centres of diversity.

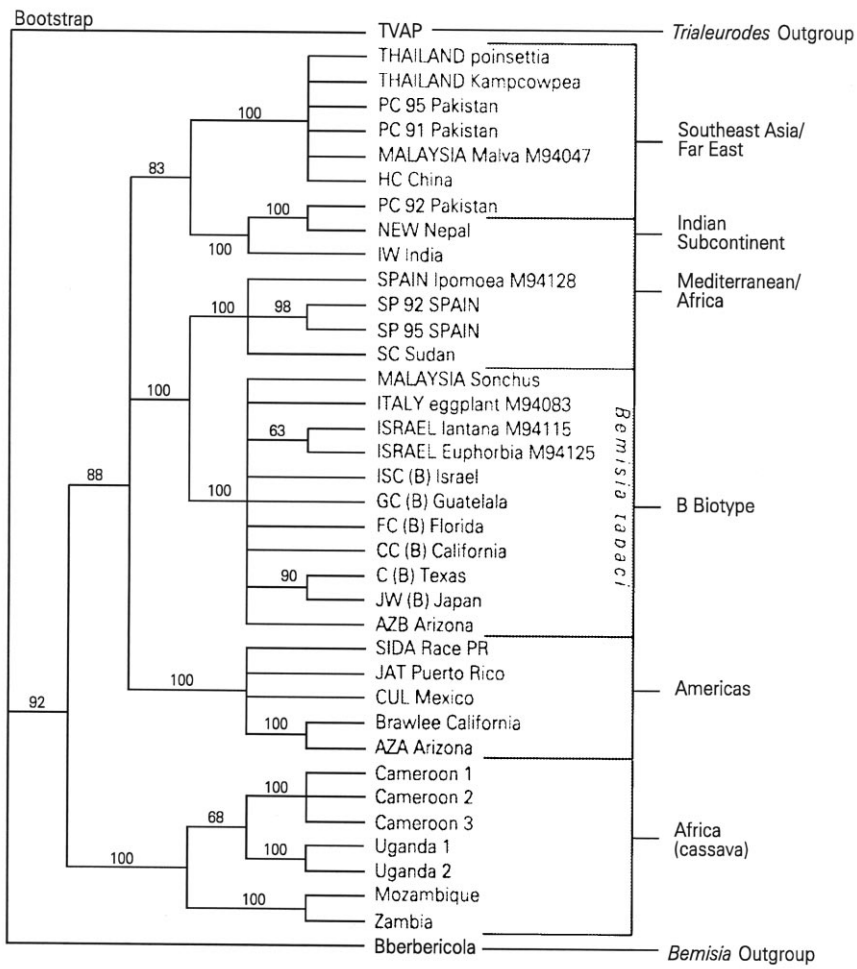

Fig. 5. Cladogram illustrating predicted relationships between representative whitefly (*B. tabaci*) mt COI sequences (708–724 nt). Tree was reconstructed following Clustal alignemnt and parsimony analysis by PAUP for 2000 bootstrap replicates with heuristic search, TBA option, and random sequence addition. This analysis yielded a single shortest tree of 1226 nt in length and nodes were collapsed when branching topology occurred less than 60% of the time. Outgroups were *T. vaporariorum* and *B. berbericola*.

Although additional data are needed on whiteflies and begomoviruses to portray an accurate contemporary map of virus–vector complexes, it is becoming clear that begomoviruses and their whitefly vector variants can be grouped according to similar geographic affiliation. Among viral gene and AA sequence trees, the *CP* AA tree was most congruent with the mt COI whitefly vector tree. This is likely to be due to the need for co-evolution between the viral capsid protein and *B. tabaci* that transmits begomoviruses in a persistent manner. Though complete details are lacking, the *CP* has been shown to harbour determinates that specify transmission of begomoviruses by this species. Although a single whitefly species transmits all begomoviruses, transmission efficiency is known to differ for particular virus–vector–host combinations (Bedford et al., 1994; Bird, 1957; Bird and Maramorosch, 1978; Brown et al., 1995b; McGrath and Harrison, 1995; Burban et al., 1992). Clearly, differences in host preference, host range, fecundity and dispersal behaviour also indite the whitefly vector in potentiating recombination and reassortment in nature, possibly exacerbating the selection of virulent and non-virulent viruses. This could occur through subtle selectivity involving the levels of ingestion, protection by chaperonins, acquisition, localization in phloem, and/or transmission efficiency itself.

Collectively, accurate virus and vector identification and sound knowledge of their distribution and importance will facilitate pest and vector control through directed and more judicious use of pesticides. Moreover, knowledge of the identity and distribution of begomoviruses in cropping systems will permit the timely development and utilization of disease-resistant cultivars to protect against the most important pathogens. Necessarily, efforts must be integrated and directed closely toward reducing crop damage and losses caused by emergent virus–vector complexes: *B. tabaci* and the geminiviruses it transmits. Clearly, studies that unravel the relationships between virus–vector–plant complexes and link phylogenetic relationships with biological, cellular, and molecular relationships also show great promise for gaining new insights into the co-evolution of organisms from distinct kingdoms.

Acknowledgements

The author would like to thank Dr. Julio Bird for endless inspirations and insights and for the many valuable discussions that made this work possible. Gratitude is also extended to the following colleagues for providing virus and/or whitefly samples used in this study: I. Bedford, J. Bird, R. Caballero, R. Gill, M.M. Roca De Doyle, A. Kirk, L. Lacey, J.P. Legg, G.W. Otim-Nape, M. Palmieri, and M.E.C. Rey, and S. Berry. Special thanks are extended to D. Krigsvold for making possible collection trips in Central America shortly after the introduction of the B biotype.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Arguello-Astorga, G.R., Guevara-Gonzalez, R.G., Herrera-Estrella, L.R., Rivera-Bustamante, R.F., 1994. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. Virology 203, 90–100.

Bedford, I.D., Briddon, R.W., Brown, J.K., Rosell, R.C., Markham, P.G., 1994. Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. Ann. Appl. Biol. 125, 311–325.

Behjatnia, S.A.A., Dry, I.B., Rezaian, M.A., 1998. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. Nucleic Acids Res. 26, 925–931.

Bird, J., 1957. A whitefly-transmitted mosaic of *Jatropha gossypifolia*. Technical Paper. Agriculture Experiment Station, Puerto Rico 22, 1–35.

Bird, J., Maramorosch, K., 1978. Viruses and virus diseases associated with whiteflies. Adv. Virus Res. 22, 55–110.

Bird, J., Sanchez, J., 1971. Whitefly-transmitted viruses in Puerto Rico. J. Agric. Univ. PR 55, 461–467.

Bisaro, D.M., 1994. Recombination in geminiviruses: mechanisms for maintaining genome size and generating genomic diversity. In: Paszkowski, J. (Ed.), Homologous Recombination and Gene Silencing in Plants. Kluwer, Dordrecht, pp. 39–60.

Bisaro, D.M., 1996. Geminivirus DNA Replication. In: DNA Replication in Eukaryotic Cells. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 833–854.

Bock, K.R., 1982. Geminivirus diseases in tropical crops. Plant Dis. 66, 266–270.

Brown, J.K., 1990. An update on the whitefly-transmitted geminiviruses in the Americas and the Caribbean Basin. FAO Plant Prot. Bull. 39, 5–23.

Brown, J.K., 1994. The status of *Bemisia tabaci* (Genn.) as a pest and vector in world agroecosystems. FAO Plant Prot. Bull. 42, 3–32.

Brown, J.K., 2000. The molecular epidemiology of begomoviruses. Chapter 13 in: Trends in Plant Virology (J.A. Khan and J. Dykstra), The Haworth Press, Inc., NY (in press).

Brown, J.K., Bird, J., 1992. Whitefly-transmitted geminiviruses in the Americas and the Caribbean Basin: past and present. Plant Dis. 76, 220–225.

Brown, J.K., Coats, S.A., Bedford, I.D., Markham, P.G., Bird, J., Frohlich, D.R., 1995a. Characterization and distribution of esterase electromorphs in the whitefly, *Bemisia tabaci* (Genn.) (Homoptera: Aleyrodidae). Biochem. Genet. 33, 205–214.

Brown, J.K., Fletcher, D., Bird, J., 1993. First report of Passiflora leaf mottle caused by a whitefly-transmitted geminivirus in Puerto Rico. Plant Dis. 77, 1264.

Brown, J.K., Frohlich, D.R., Rosell, R.C., 1995b. The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex? Ann. Rev. Entomol. 40, 511–534.

Brown, J.K., Orstrow, K.M., Idris, A.M., Stenger, D.C., 2000a. Chino del tomate virus: relationships to other begomoviruses and the identification of A component variants that affect symptom expression. Phytopathology 90, 546–552.

Brown, J.K., Torres-Jerez, I., Idris, A.M., Banks, G.K., Wyatt, S.D., 2000. The core region of the coat protein gene is highly useful for establishing the provisional identification and classification of begomoviruses. Arch. Virol., submitted for publication

Brown, J.K., Idris, A.M., Olsen, M., Miller, M.E., Isakeit, T., Anciso, J., 2000b. *Cucurbit leaf curl virus*, a new whitefly-transmitted geminivirus in Arizona, Texas and Mexico. Plant Dis. 84, 809.

Burban, C., Fishpool, L.D.C, Fauquet, C., Fargette, D., Thouvenel, J.-C., 1992. Host-associated biotypes within West African populations of the whitefly *Bemisia tabaci* (Genn.), Homoptera: Aleyrodidae. J. Appl. Entomol. 113, 416–423.

Byrne, D.N., Bellows, T.S., Jr., Parrella, M.P., 1990. Whiteflies in agricultural systems. In: Whiteflies: their bionomics, pest status, and management. Intercept, Andover, Hants, UK, pp. 227–261.

Campbell, B.C., Steffen-Campbell, J.D., Gill, R.J., 1996. Origin and radiation of whiteflies: an initial molecular phylogenetic assessment. In: Gerling, D., Mayer, R.T. (Eds.), *Bemisia*: 1995 Taxonomy, Biology, Damage, Control, and Management. Intercept, Andover, Hants, UK, pp. 29–51.

Cock, M.J.W., 1993. *Bemisia tabaci*, an update 1986–1992 on the cotton whitefly with an annotated bibliography. CAB IIBC, Silwood Park, UK, p. pp 78.

Costa, A.S., 1976. Whitefly-transmitted plant diseases. Annu. Rev. Phytopathol. 14, 429–449.

Costa, A.S., Russell, L.M., 1975. Failure of *Bemisia tabaci* to breed on cassava plants in Brazil (Homoptera: Aleyrodidae). Cien. e Cult. 27, 388–390.

Costa, H.S., Brown, J.K., 1991. Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci* (Genn.) and the association of one population with silverleaf symptom induction. Entomologia Experimentalis et Applicata 61, 211–219.

Costa, H.S., Brown, J.K., Sivasupramanium, S., Bird, J., 1993. Regional distribution insecticide resistance and reciprocal crosses between the type A and type B of *Bemisia tabaci*. Insect Sci. Appl. 14, 255–266.

De Barro, P.J., Driver, F., 1997. Use of RAPD PCR to distinguish the B biotype from other biotypes of *B. tabaci* (Gennadius) (Hemiptera: Aleyrodidae). Aust. J. Entomol. 36, 149–152.

De Barro, P.J., Driver, F., Trueman, J.W.H, Curran, J., 2000. Phylogenetic relationships of world populations of *Bemisia tabaci* (Gennadius) using ribosomal ITS1. Mol. Phylogenet. Evol. 16, 29–36.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Duffus, J.E., 1987. Whitefly-transmission of plant viruses. In: Harris, K.F. (Ed.), Current Topics in Vector Research, vol. 4. Springer, New York, pp. 73–91.

Eagle, P.A., Orozco, B.M., Hanley-Bowdoin, L., 1994. A DNA sequence required for geminivirus Replication also mediates transcriptional regulation. The Plant Cell 6, 1157–1170.

Fontes, E.P.B., Gladfelter, H.J., Schaffer, R.L., Petty, I.T.D., Hanley-Bowdoin, L., 1994. Geminivirus replication origins have a modular organization. The Plant Cell 6, 405–416.

Frohlich, D., Torres-Jerez, I., Bedford, I.D., Markham, P.G., Brown, J.K, 1999. A phylogeographic analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. Mol. Ecol. 8, 1593–1602.

Frohlich, D.R, Brown, J.K., Bedford, I., Markham, P.G., 1996. Mitochondrial 16S ribosomal subunit as a molecular marker in *Bemisia*, and implications for population variability. In: *Bemisia* 1995: Taxonomy, Biology, Damage Control, and Management. Intercept, Andover, Hants, UK, pp. 143–145.

Fuchs, M.A., Buta, C., 1997. The role of peptide modules in protein evolution. Biophys. Chem. 66, 203–210.

Gawel, N.J., Bartlett, A.C., 1993. Characterization of differences between whiteflies using RAPD-PCR. Insect Mol. Biol. 2, 33–38.

Gill, R.J., 1992a. The morphology of whiteflies. In: Gerling, D. (Ed.), Whiteflies: their Bionomics, Pest Status, and Management. Intercept, Andover, Hants, UK, pp. 13–46.

Gill, R.J., 1992b. A review of the sweetpotato whitefly in southern California. Pan-Pacific Entomologist 68, 144–152.

Gladfelter, H.J., Eagle, P.A., Fontes, E.P.B., Batts, L., Hanley-Bowdoin, L., 1997. Two domains of the AL1 protein mediate geminivirus origin recognition. Virology 239, 186–197.

Goodman, R.L., 1981. Geminiviruses. J. Gen. Virol. 54, 9–21.

Guirao, P., Beitia, F., Cenis, J.L., 1997. Biotype determination of Spanish populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae). Bull. Entomol. Res. 87, 587–593.

Hanley-Bowdoin, L., Settlage, S.B., Orozco, B.M., Nagar, S., Robertson, D., 1999. Geminiviruses: models for plant DNA replication, transcription and cell cycle regulation. Crit. Rev. Plant Sci. 18, 71–106.

Hansen, S.F., Hoogstraten, R.A., Ahlquist, P., Gilbertson, R.L., Russell, D.R., Maxwell, D.P., 1995. Mutational analysis of a putative NTP-binding domain in the replication-associated protein (AC1) of bean golden mosaic geminivirus. Virology 211, 1–9.

Harrison, B.D., 1985. Advances in geminivirus research. Ann. Rev. Phytopathol. 23, 55–82.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., Gronenborn, B., 1998. Determination of the origin cleavage and joining domain of geminivirus Rep proteins. Nucleic Acids Res. 23, 910–916.

Hill, J.E., Strandberg, O., Hiebert, E., Lazarowitz, S.G., 1998. Asymmetric infectivity of pseudorecombinants of cabbage leaf curl virus and squash leaf curl virus: implications for bipartite geminivirus evolution and movement. Virology 250, 283–292.

Hou, Y.M., Gilbertson, R.L., 1996. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. J. Virol. 70, 5430–5436.

Idris, A.M., Brown, J.K., 1998. Sinaloa tomato leaf curl geminivirus: biological and molecular evidence for a new subgroup III virus. Phytopathology 88, 648–657.

Idris, A.M., Bird, J., Brown, J.K., 1999. First report of a bean-infecting begomovirus from *Macroptilium lathyroides* in Puerto Rico that is distinct from bean golden mosaic virus. Plant Dis. 83, 1071.

Idris, A.M., Brown, J.K., 2000. Identification of new, monopartite begomovirus associated with leaf curl disease of cotton in Gezira, Sudan. Plant Dis. 84, 809.

Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eukaryotes and archaebacteria. Nucleic Acids Res. 20, 3279–3285.

Ingham, D.J., Pascal, E., Lazarowitz, S.G., 1995. Both bipartite movement proteins define viral host range but only BL1 determines pathogenicity. Virology 207, 191–204.

Jeffrey, J.L., Pooma, W., Petty, I.T.D., 1996. Genetic requirements for local and systemic movement of tomato golden mosaic virus in infected plants. Virology 223, 208–218.

Jupin, I., Hericourt, F., Benz, B., Gronenborn, B., 1995. DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the REP protein. FEBS Lett. 362, 116–120.

Kirk, A.A, Lacey, L.A., Brown, J.K., Ciomperlik, M.A., Goolsby, J.A., Vacek, D.C., Wendel, L.E., Napompeth, B., 2000. Variability within the *Bemisia tabaci* complex (Homoptera: Aleyrodidae) and its natural enemies: successful biological control of type B in the USA. Bull. Entom. Res., in press.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., Gronenborn, B., 1995a. Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. Biochimie 77, 765–773.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J., Gronenborn, B., 1995b. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. Proc. Natl. Acad. Sci. USA 92, 3879–3883.

Lazarowitz, S.G., 1991. Molecular characterization of two bipartite geminiviruses causing squash leaf curl disease: role of viral replication and movement functions determining host range. Virology 180, 70–80.

Lazarowitz, S.G., 1992. Geminiviruses: genome structure and gene function. Crit. Rev. Plant Sci. 11, 327–349.

Lazarowitz, S.G., Wu, L.C., Rogers, S.G., Elmer, J.S., 1992. Sequence specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. The Plant Cell 4, 799–809.

Mayo, M.A., Pringle, C.R., 1998. Virus Taxonomy – 1997. J. Gen. Virol. 79, 649–657.

McGrath, P.F., Harrison, B.D., 1995. Transmission of tomato yellow leaf curl geminivirus by *Bemisia tabaci*: effects of virus isolate and vector biotypes. Ann. Appl. Biol. 126, 307–316.

Mound, L.A., 1983. Biology and identity of whitefly vectors of plant pathogens. In: Plumb, R.T., Thresh, J.M. (Eds.), Plant Virus Epidemiol. Blackwell, Oxford, UK, pp. 305–313.

Mound, L.A., Halsey, S.H. (Eds.), 1978. Whitefly of the world: a systematic catalogue of the Aleyrodidae (Homoptera) with host plant and natural enemy data. British Museum (Natural History), London, UK. Wiley, Chichester.

Muniyappa, V., 1980. Whiteflies. In: Harris, K.F., Maramorosch, K. (Eds.), Vectors of Plant Pathogens. Academic Press, New York, pp. 39–85.

Nakhla, M.K., Maxwell, D.P., Martinez, R.T., Carvalho, M.G., Gilbertson, R.L., 1994. Widespread occurrence of the eastern Mediterranean strain of tomato yellow leaf curl geminivirus in tomatoes in the Dominican Republic. Plant Dis. 78, 926.

Noueiry, A.O., Lucas, W.J., Gilbertson, R.L., 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmatal transport. Cell 76, 925–932.

Orozco, B.M., Gladfelter, H.J., Settlage, S.B., Eagle, P.A., Gentry, R.N., Hanley-Bowdoin, L., 1998. Multiple cis elements contribute to geminivirus origin function. Virology 242, 346–356.

Orozco, B.M., Miller, A.B., Settlage, S.B., Hanley-Bowdoin, L., 1997. Functional domains of a geminivirus replication protein. J. Biol. Chem. 272, 9840–9846.

Padidam, M., Beachy, R.N., Fauquet, C.M., 1995. Classification and identification of geminiviruses using sequence comparisons. J. Gen. Virol. 76, 249–263.

Pascal, E., Sanderfoot, A.A., Ward, B.M., Medville, R., Turgeon, R., Lazarowitz, S.G., 1994. The geminivirus BR1 movement protein binds single-stranded DNA and localizes to the cell nucleus. The Plant Cell 6, 995–1006.

Paximadis, M., Idris, A.M., Torres-Jerez, I., Villarreal, I., Rey, M.E.C., Brown, J.K., 1999. Characterization of geminiviruses of tobacco in the Old and New World. Arch. Virol. 144, 703–717.

Perring, T.M., Cooper, A.D., Russell, R.J., Farrar, C.A., Bellows, T.S., Jr., 1993. Identification of a whitefly species by genomic and behavioral studies. Science 259, 74–77.

Petty, I.T.D., Miller, C.G., Meade-Hash, T.J., Schaffer, R.L., 1995. Complementable and non-complementable host adaptation defects in bipartite geminiviruses. Virology 212, 263–267.

Polston, J.E., Anderson, P.K., 1997. The emergence of whitefly-transmitted geminiviruses in tomato in the Western hemisphere. Plant Dis. 81, 1358–1369.

Polston, J.E., Bois, D., Serra, C.A., Concepcion, S., 1994. First report of a tomato yellow leaf curl-like geminivirus in the Western Hemisphere. Plant Dis. 78, 831.

Rey M.E.C., Berry, S., Banks, G.K., Markham, P.G., Brown, J.K, 2000. A study of whitefly populations on cassava in southern Africa. In: Proc. Int. Symp. on Tropical Root and Tuber Crops, Thiruvananthapuram, India, January 19–22, 2000.

Rigden, J.E., Dry, I.B., Krake, L.R., Rezaian, M.A., 1996. Plant virus DNA Replication processes in *Agrobacterium*: insight into the origins of geminiviruses? Proc. Natl. Acad. Sci. 93, 10280–10284.

Rosell, R.C., Bedford, I.D., Frohlich, D.R., Gill, R.J., Brown, J.K., Markham, P.G., 1997. Analysis of morphological variation in distinct populations of *Bemisia tabaci* (Homoptera: Aleyrodidae). Ann. Entomol. Soc. Am. 90, 575–589.

Russell, L.M., 1957. Synonyms of *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae). Bull. Brooklyn Entomol. Soc. 52, 122–123.

Rybicki, E.P., 1994. A phylogenetic and evolutionary justification for three genera of Geminiviridae. Arch. Virol. 139, 49–77.

Rybicki, E.P., 1998. A proposal for naming geminiviruses; a reply by the *Geminiviridae* Study Group Chair. Arch. Virol. 143, 421–424.

Sanderfoot, A.A., Lazarowitz, S.G., 1995. Cooperation in viral movement: the geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. The Plant Cell 7, 1185–1194.

Sanderfoot, A.A., Lazarowitz, S.G., 1996. Getting it all together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. Trends Cell Biol. 6, 353–358.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Stanley, J., 1995. Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling circle replication. Virology 206, 707–712.

Swofford, D., 1991. PAUP user's manual. Version 3.1.1. Illinois Natural History Survey, Champaign, IL.

Traboulsi, R., 1994. *Bemisia tabaci*: a report on its pest status with particular reference to the Near East. FAO Plant Prot. Bull. 42, 33–58.

Varma, P.M., 1963. Transmission of plant viruses by whiteflies. Bull. Natl. Inst. Sci. 24, 11–33.

Wool, D., Gerling, D., Belloti, A.C, Morales, F.J., 1993. Esterase electrophoretic variation in *Bemisia tabaci* (Genn.) (Homoptera: Aleyrodidae) among host plants and localities in Israel. J. Appl. Entomol. 115, 185–196.

Wyatt, S.D., Brown, J.K., 1996. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. Phytopathology 86, 1288–1293.

Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G.W., Robinson, D., Harrison, B.D., 1997. Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. J. Gen. Virol. 78, 2101–2111.

Zhou, X, Liu, Y, Robinson, D., Harrison, B.D., 1998. Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. J. Gen. Virol. 79, 915–923.