Cotton leaf curl disease – an emerging threat to cotton production worldwide

M. Naeem Sattar,1 Anders Kvarnheden,1 Muhammad Saeed2 and Rob W. Briddon2

1Department of Plant Biology and Forest Genetics, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Box 7080, SE-750 07 Uppsala, Sweden
2National Institute for Biotechnology and Genetic Engineering, PO Box 577, Jhang Road, Faisalabad, Pakistan

Cotton leaf curl disease (CLCuD) is a serious disease of cotton which has characteristic symptoms, the most unusual of which is the formation of leaf-like enations on the undersides of leaves. The disease is caused by whitefly-transmitted geminiviruses (family Geminiviridae, genus Begomovirus) in association with specific, symptom-modulating satellites (betasatellites) and an evolutionarily distinct group of satellite-like molecules known as alphasatellites. CLCuD occurs across Africa as well as in Pakistan and north-western India. Over the past 25 years, Pakistan and India have experienced two epidemics of the disease, the most recent of which involved a virus and satellite that are resistance breaking. Loss of this conventional host–plant resistance, which saved the cotton growers from ruin in the late 1990s, leaves farmers with only relatively poor host plant tolerance to counter the extensive losses the disease causes. There has always been the fear that CLCuD could spread from the relatively limited geographical range it encompasses at present to other cotton-growing areas of the world where, although the disease is not present, the environmental conditions are suitable for its establishment and the whitefly vector occurs. Unfortunately recent events have shown this fear to be well founded, with CLCuD making its first appearance in China. Here, we outline recent advances made in understanding the molecular biology of the components of the disease complex, their interactions with host plants, as well as efforts being made to control CLCuD.

Introduction

Cotton leaf curl disease (CLCuD) is a serious disorder of several plant species in the family Malvaceae, the most important of which is cotton (Genus: Gossypium L.). The disease occurs across Africa and southern Asia. Plants affected by the disease exhibit very unusual symptoms, consisting of vein swelling, upward or downward cupping of the leaves, and the formation of enations on the main veins on the undersides of leaves (Fig. 1). Frequently the enations develop into cup-shaped, leaf-like structures which can become as large as the leaf from which they emerge. Unusually CLCuD-affected cotton plants appear greener than non-infected plants due to the proliferation of chloroplast-containing tissues. However, symptoms are variable with cotton variety and, particularly, the age of the plant at infection. Late season infection frequently leads to mild symptoms and little yield reduction. Plants infected soon after germination are usually severely stunted, with tightly rolled leaves, and produce no harvestable lint.

Cotton has long been a major source of food, feed and fibre across the world. The use of cotton fibres to make fabrics goes back at least 7000 years (Cantrell, 2005; Sunilkumar et al., 2006). Globally ~32.6 million hectares are devoted to cotton cultivation with production estimated at 27.6 million tons for 2011/2012 (Anonymous, 2011). The genus Gossypium encompasses ~51 species. Of the cultivated species, Gossypium hirsutum L. and Gossypium barbadense L. are allotetraploid, whilst Gossypium arboreum L. and Gossypium herbaceum L. have diploid genomes (Wendel & Cronn, 2003). The tetraploid species G. hirsutum (often referred to as upland cotton) and G. barbadense L. are allotetraploid, whilst Gossypium arboreum L. and Gossypium herbaceum L. have diploid genomes (Wendel & Cronn, 2003). The tetraploid species G. hirsutum (often referred to as upland cotton) and G. barbadense (known as Pima/Egyptian cotton), have their centre of origin in Central America and northern Peru, respectively. G. hirsutum is the most widely cultivated species due to the quality of the fibre it produces, which is long with a thick secondary wall, giving it strength. The diploid species G. arboreum and G. herbaceum are native to the Old World (OW) but the fibre they produce is not valued as highly by processors. However, G. arboreum, although producing a lower quality fibre, is adapted to hot dry environments and is immune to CLCuD, and is thus still an important commercial species.
There have been a number of reviews detailing the history, geographical distribution and elucidation of the aetiology of CLCuD and this information will not be recapitulated here (Briddon & Markham, 2000; Mansoor et al., 2011; Farooq et al., 2011). The CLCuD system was amongst the first in which single-stranded DNA satellites associated with geminiviruses were identified (Briddon & Stanley, 2006) and is one of the most intensively studied. There have been significant advances in our understanding of the interactions between the components of the disease complex and of the interactions of the components of the complex with host plants. This information, as well as the evidence for the complex spreading from its origin in Pakistan/India will be dealt with in this review.

Components of the CLCuD begomovirus complex
CLCuD-associated begomoviruses
Viruses of the family Geminiviridae have small, ssDNA genomes that are encapsidated in characteristic twinned quasi-icosahedral particles. Geminiviruses are assigned to one of the four genera (Topocuvirus, Curtovirus, Mastrevirus or Begomovirus) based upon genome arrangement, insect vector and sequence relatedness (Brown et al., 2012). Viruses of the genus Begomovirus, which are transmitted by the ubiquitous whitefly Bemisia tabaci, are the most numerous and economically the most destructive. Begomoviruses originating from the New World (NW) typically have genomes consisting of two components, known as DNA A and DNA B (Fig. 2), both of which are required for virus infectivity (Stanley, 1983). However, in the OW, although there are a small number of bipartite begomoviruses, the majority have genomes consisting of only a single genomic component, which is a homologue of the DNA A of the bipartite viruses. Of these, a small number are truly monopartite; their single component induces disease in plants in the field, such as Tomato yellow leaf curl Sardinia virus (TYLCSV) (Kheyr-Pour et al., 1991). The majority of monopartite begomoviruses associate with a recently identified class of satellites (betasatellites) and satellite-like molecules (alphasatellites; Briddon & Stanley, 2006). The structures of the genomes, genomic components and satellites associated with begomoviruses are summarized in Fig. 2.

In the NW, malvaceous species (including cotton) are affected by bipartite begomoviruses. The virus affecting cotton identified in the southern United States is Cotton leaf crumple virus (CLCrV; Idris & Brown, 2004). In most years this virus does not cause significant losses for cotton production, as it usually does not infect plants until late in the season (Briddon & Markham, 2000). In common with all NW begomoviruses, CLCrV is distinct from the begomoviruses in the OW (Fig. 3), likely because they have been evolving independently for some considerable time (Lefevre et al., 2011).

In the OW two distinct CLCuD complexes are recognized at this time: the African and Asian complexes. Although CLCuD was first reported from Africa in 1912 (Farquharson, 1912), the causative agent was not identified until much later. Only a
single virus, Cotton leaf curl Gezira virus (CLCuGeV; Idris & Brown, 2002), has been identified in cotton in Africa; however, few cotton samples have been analysed and the actual diversity present may be much larger than we presently realize. CLCuGeV is a geographically widespread species occurring from central Africa to Jordan and infects a number of plant species including cotton, okra, hollyhock and Sida spp. (Tahir et al., 2011).

The situation on the Indian subcontinent is more complex than that in Africa. During the epidemic of CLCuD in Pakistan and north-western India, at least six distinct begomovirus species were identified in cotton and many plants were found to be infected with more than one species. The species identified were Cotton leaf curl Alabad virus (CLCuAIV), Cotton leaf curl Kokhran virus (CLCuKoV), Cotton leaf curl Multan virus (CLCuMuV), Cotton leaf curl Rajasthan virus (CLCuRaV), Papaya leaf curl virus (PaLCuV) and Tomato leaf curl Bangalore virus (ToLCBaV; note that geminiviruses are named after the first host they are identified in and the name does not therefore necessarily indicate the host range or host preference of a particular species) (Kirthi et al., 2004; Zhou et al., 1998; Fig. 3). However, of these species Koch’s postulates have only been satisfied (in cotton) for CLCuMuV, CLCuKoV and PaLCuV and in each case, the virus required the presence of a specific betasatellite [Cotton leaf curl Multan betasatellite (CLCuMuB)] to infect cotton (Briddon et al., 2001; Mansoor et al., 2003b). Additionally Cotton leaf curl Bangalore virus (CLCuBaV), a further species associated with CLCuD, was identified in southern India (Chowda Reddy et al., 2005). The virus was

---

**Fig. 2.** Genome organization of begomoviruses. Begomoviruses have either bipartite or monopartite genomes. Those native to the NW have bipartite genomes with components known as DNA A and DNA B. The bipartite begomoviruses in the NW differ from those in the OW in lacking the V2 gene. Although some begomoviruses in the OW have bipartite genomes, the majority are monopartite, lacking the DNA B component. Most monopartite begomoviruses are associated with the satellites known as betasatellites and in many cases, these begomovirus–betasatellite complexes are additionally associated with the satellite-like alphasatellites. The proteins encoded by the components are indicated as the replication associated protein (Rep), the transcriptional activator protein (TrAP; the homologue of which for some monopartite viruses is known as C2, since it does not transactivate late, virion-sense-encoded genes), the replication enhancer protein (REn), the C4 protein (C4), the coat protein (CP) and the V2 protein (V2) on the genomes of monopartite viruses and the DNA A component of bipartite viruses. The DNA B component encodes the nuclear shuttle protein (NSP) and the movement protein (MP). A sequence conserved between the DNA A and DNA B components of bipartite begomoviruses is known as the conserved region (CR). The betasatellites have a single gene ([jIC1] in the complementary-sense and a region of sequence highly conserved between all betasatellites [known as the satellite conserved region (SCR)] and a region of sequence rich in adenine (A-rich). The alphasatellites have a single large protein (Rep) and contain an A-rich sequence. For each component the conserved hairpin structure, containing the nonanucleotide sequence TAATATTAC (TAGTATTAC for most alphasatellites) within the loop structure, is shown at position zero.
shown to be associated with *Kenaf* leaf curl betasatellite (KeLCuB), a betasatellite that occurs in another malaceous fibre crop kenaf (*Hibiscus canabinus* L.; Fig. 4; Paul et al., 2008). CLCuD is not a problem in southern India. CLCuBaV and KeLCuB, were not part of the epidemic that occurred in the north, and have not been identified in cotton since, likely indicating that they are infrequent pathogens of cotton.

Following resistance breaking in cotton in 2001 only a single begomovirus [*Cotton leaf curl Burewala virus* (CLCuBaV)] was found across the Punjab, Pakistan (Fig. 5; Amrao et al., 2010b). Although, for the most part, the Sindh province of Pakistan has been free of CLCuD, CLCuD has become a problem there since 2004, although not on the scale of the rest of the country (Mansoor et al., 2006). Recent characterization of the viruses involved has shown a more diverse population of begomoviruses than is present in Punjab to the north at this time (Amrao et al., 2010a). As well as CLCuKoV, a virus which was part of the CLCuD epidemic in the 1990s, CLCuGeV (as will be discussed later) and a new species, *Cotton leaf curl Shahdadpur virus* (CLCuShV) were identified. Both CLCuBaV and CLCuShV have recombinant genomes, with sequences derived from CLCuKoV and CLCuMuV.

During both the ‘Multan’ and ‘Burewala’ CLCuD epidemics, the disease spread eastwards out of Pakistan into the cotton-growing states of north-western India. Many of the virus species (CLCuMuV, CLCuKoV and CLCuBaV) identified in Pakistan were subsequently identified in India (Rajagopalan et al., 2012; Zaffalon et al., 2012). However, some species were identified in India that have not appeared in cotton in Pakistan. CLCuRaV has been identified extensively in cotton in India but, except for being identified in a collection of exotic (not cultivated) cotton species in Multan (Nawaz-ul-Rehman et al., 2010) and tomato in Faisalabad (Shahid et al., 2007), it has not been identified in cotton in Pakistan. This species may thus be one of very few cases where a cotton begomovirus has spread out of India into Pakistan. Similarly ToLCBaV has only been identified in cotton in India and on only one occasion, which may suggest that this virus was an accidental infection of cotton (possibly maintained by one of the other cotton-infecting viruses), rather than being a significant pathogen of cotton. This may also be the case for the one report of the bipartite begomovirus *Tomato leaf curl New Delhi virus* (ToLCNDV; GenBank accession no. EF063145) in cotton in India. The most recent study of begomovirus diversity in cotton, after resistance breaking, in north-western India has shown CLCuBuV and CLCuRaV to be present (Rajagopalan et al., 2012), which differs from the situation in Pakistan. One possible explanation for this difference is that the uptake of resistant varieties was not as extensive in India as in Pakistan due to the patchy nature of the disease. For farmers it was always a gamble about whether to grow the lower yielding resistant varieties (in case CLCuD incidence was high) or gamble on low disease incidence and grow high yielding, but susceptible, varieties.

**Betasatellites**

Although only first identified as recently as 1999–2000 (Briddon et al., 2001; Saunders et al., 2000), our understanding of betasatellites has progressed at a rapid pace. The betasatellites are a diverse class of ssDNA molecules of approximately half the size of their helper begomoviruses (~1350nt) that share no sequence homology with their helper viruses, other than the presence of a potential stem–loop structure containing the ubiquitous nonanucleotide sequence TAATAATAC (Briddon et al., 2003; Fig. 2). The stem–loop structure occurs within an ~100 nt sequence, known as the satellite conserved region (SCR), which is highly conserved between all betasatellites.

---

*Fig. 3.* Neighbour-joining phylogenetic dendrogram based upon an alignment of the full-length nucleotide sequences of begomoviruses infecting cotton and other selected begomoviruses. The origins of begomoviruses are shown along the right hand side (either originating from the NW or the OW). The begomoviruses associated with cotton are *Cotton leaf curl Alabad virus* (CLCuAlV), *Cotton leaf curl Bangalore virus* (CLCuBaV), *Cotton leaf curl Burewala virus* (CLCuBuV), *Cotton leaf curl crumple virus* (CLCrV), *Cotton leaf curl Gezira virus* (CLCuGeV; the isolates originating from Pakistan are highlighted (PK)), *Cotton leaf curl Khoran virus* (CLCuKoV), *Cotton leaf curl Multan virus* (CLCuMuV; which consist of multiple strains, of which two, the Faisalabad (Fai) and Hisar (His) strains, are shown), *Cotton leaf curl Rajasthan virus* (CLCuRaV), *Cotton leaf curl Shahdadpur virus* (CLCuShV), *Papaya leaf curl virus* (PaLCuV), *Tomato leaf curl Bangalore virus* (ToLCBaV). The isolates of CLCuMuV originating from China are highlighted (CN). For virus species associated with cotton diseases (although not necessarily isolated from cotton) the isolates are indicated as originating from Burkina Faso (BF), Cameroon (CM), China (CN), Egypt (EG), India (IN), Mali (MA), Niger (NE), Pakistan (PK), Sudan (SD) and the United States (US). Plants from which isolates were obtained are indicated as *Abelmoschus esculentus* (AE; okra), *Althea rosea* (AR; hollyhock), *Carica papaya* (CP; papaya), *Gossypium barbadense* (GB), *G. hirsutum* (GH) and *Hibiscus rosa-sinensis* (HR). The other viruses are *African cassava mosaic virus* (ACMV), *Ageratum leaf curl virus* (ALCuV), *Cabbage leaf curl virus* (CbLCuV), *Chili leaf curl virus* (ChLCV), *Indian cassava mosaic virus* (ICMV), *Pepper leaf curl Bangladesh virus* (PepLBDV), *Pepper leaf curl Lahore virus* (PepLCLaV), *Tomato leaf curl New Delhi virus* (ToLCNDV), *Tomato yellow leaf curl China virus* (TYLCCNV) and *Tomato yellow leaf curl virus* (TYLCV). Numbers at nodes indicate percentage bootstrap confidence scores (1000 replicates). Vertical distances are arbitrary, horizontal distances are proportional to calculated mutation distances. The tree was rooted on the distantly related *Tomato pseudo-curl top virus* (TPCtV) as outgroup.
Fig. 4. Neighbour-joining phylogenetic dendrogram based upon an alignment of the full-length nucleotide sequences of betasatellites infecting cotton and selected other betasatellites. The betasatellite species used were *Ageratum* leaf curl Cameroon betasatellite (ALCCMB), *Ageratum* yellow leaf curl betasatellite (AYLCuB), *Ageratum* yellow vein Sri Lanka betasatellite (AYVSLB), *Alternanthera* yellow vein betasatellite (AYVB), Bean leaf curl China betasatellite (BeLCNNB), Bhendi yellow vein betasatellite (BYVB), Chili leaf curl betasatellite (ChlCB), Cotton leaf curl Gezira betasatellite (CLCuGeB), Cotton leaf curl Multan betasatellite (CLCuMuB), Croton yellow vein betasatellite (CroYVMN), *Emilia* yellow vein betasatellite (EmYVB), *Erectites* yellow mosaic betasatellite (ErYMB), *Eupatorium* yellow vein betasatellite (EpYVB), Honeysuckle yellow vein betasatellite (HYVB), Honeysuckle yellow vein Kochi betasatellite (HYVKoB), *Lindernia anagallis* yellow vein betasatellite (LaYVB), *Kenaf* leaf curl betasatellite (KeLCuB), *Leucas zeylanica* yellow vein betasatellite (LeZYVB), *Luffa* leaf distortion betasatellite (LuLDB), *Malvastrum* leaf curl betasatellite (MaLCuB), *Malvastrum* yellow vein Yunnan betasatellite (MaYVYNB), Okra leaf curl betasatellite (OCLuB), Papaya leaf curl betasatellite (PaLCuB), Radish leaf curl betasatellite (RaLCB), *Sida* leaf curl betasatellite (SiLCuB), *Sida* yellow vein China betasatellite (SiYVCNB), *Sida* yellow vein betasatellite (SiYVB), *Sida* yellow vein Vietnam betasatellite (SiYVNB), *Sugesbeckia* yellow vein betasatellite (SgYVB), Tobacco leaf curl betasatellite (TbLCB), Tobacco curly shoot betasatellite (TbCSB), Tomato leaf curl Bangladesh betasatellite (ToLCBD), Tomato leaf curl betasatellite (ToLCB), Tomato leaf curl China betasatellite (ToLCNNB), Tomato leaf curl Java betasatellite (ToLCJaB), Tomato leaf curl Joydebpur betasatellite (ToLJCuB), Tomato leaf curl Karnata betasatellite (ToLCuB), Tomato leaf curl Laos betasatellite (ToLCAB), Tomato leaf curl Maharasstra betasatellite (ToLMAu), Tomato leaf curl Patna betasatellite (ToLCPaB), Tomato leaf curl Philippines betasatellite (ToLCIP), Tomato leaf curl Ranchi betasatellite (ToLCRaB), Tomato yellow dwarf betasatellite (ToYDB), Tomato yellow leaf curl China betasatellite (TYLCuB), Tomato yellow leaf curl Thailand betasatellite (TLYCThB), Tomato yellow leaf curl Vietnam betasatellite (TYLCVN), Tomato yellow leaf curl Yunnan betasatellite (TYLChynB) and *Zinnia* leaf curl betasatellite (ZlCB). The betasatellites that affect cotton are shown as white text on a red background (note that on one occasion KeLCuB has been isolated from cotton). Betasatellites are indicated as originating from Bangladesh (BD), Burkina Faso (BF), Cameroon (CM), China (CN), Egypt (EG), India (IN), Indonesia (ID), Japan (JP), Laos (LA), Oman (OM), Mali (MA), Malaysia (MY), Niger (NE), Pakistan (PK), Sri Lanka (LK), Singapore (SG), Sudan (SD), Thailand (TH), Vietnam (VN). Numbers at significant nodes indicate percentage bootstrap confidence scores (1000 replicates).
Additionally the betasatellites contain a sequence rich in adenine (A-rich) and a single coding sequence (the \( \beta \)C1 gene) in the complementary sense.

Betasatellites depend upon their helper viruses for replication, movement in plants and transmission between plants, presumably by trans-encapsidation in the helper virus' coat protein. However, although betasatellites require the helper virus for replication (thus presumably the helper virus-encoded Rep), they do not contain the iterons of their helper viruses. Although it was initially assumed that the SCR would be involved in mediating betasatellite–helper virus Rep interactions, since the SCR encompasses a position analogous to that containing the iterons of geminiviruses, this has not proven to be the case. Deletion analysis and naturally occurring mutants suggest that, in fact, it is the sequence between the A-rich region and the SCR which may be involved in Rep binding (Nawaz-ul-Rehman et al., 2009; Saunders et al., 2008). This region is highly variable, contains sequences which resemble helper virus iterons and is possibly an adaptation allowing betasatellites to rapidly adapt to distinct begomoviruses (distinct Rep recognition sequences).

All functions thus far ascribed to betasatellites are mediated by the \( \beta \)C1 protein. For CLCuMuB, the \( \beta \)C1 protein has been shown to be a pathogenicity determinant (determines the symptoms of the infection; Qazi et al., 2007; Saeed et al., 2005), a suppressor of PTGS (overcoming host plant defences; Amin et al., 2011a) and may be involved in virus movement in planta (Saheed et al., 2007), modulating the levels of developmental microRNAs (Amin et al., 2011c) and upregulating viral DNA levels in planta (Briddon et al., 2001). For other betasatellites \( \beta \)C1 has been shown to bind DNA/RNA (Cui et al., 2005) and interact with a variety of host factors (Eini et al., 2009; Yang et al., 2008), interact with CP of the helper virus (Kumar et al., 2006) as well as being multimeric (Cheng et al., 2011) and suppressing jasmonic acid responses in plants allowing a quicker build-up of whitefly vectors (Zhang et al., 2012).

**Alphasatellites**

The third component of begomovirus–betasatellite complexes are the alphasatellites. Alphasatellites are not true satellites, as they are capable of autonomous replication, and are thus best described as satellite-like. These molecules are approximately half the size (~1400 nt) of the genomes of their helper begomoviruses, although they are somewhat larger than betasatellites (Briddon et al., 2004). They have a highly conserved structure consisting of a single large gene, in the virion-sense, which encodes a Rep protein, an A-rich sequence and encompasses a predicted hairpin structure with a nonanucleotide sequence (TAGTATTAC) forming part of the loop (Briddon et al., 2004; Mansoor et al., 1999).

The initial analyses, conducted soon after the discovery of the first alphasatellite in CLCuD-affected cotton, suggested that alphasatellites were genetically all very closely related (Briddon et al., 2004). However, accumulating evidence and a recent study conducted on CLCuD-affected cotton species from Pakistan, have shown that the diversity is much greater than previously assumed (Nawaz-ul-Rehman et al., 2012; Fig. 6). Although the majority of alphasatellites have been identified in association with monopartite begomoviruses originating from the OW, and always in association with a betasatellite, recent studies have identified alphasatellites in the NW in association with bipartite begomoviruses and in the absence of betasatellites (Paprotka et al., 2010; Romay et al., 2010) and in the OW in association with monopartite begomoviruses in the absence of betasatellites (Leke et al., 2011).

The precise benefits to a begomovirus–betasatellite complex of the presence of an alphasatellite remain unclear. Initial studies detected only a slight reduction in the levels of viral DNA in plants infected with begomovirus–betasatellites in the presence of an alphasatellite, leading to the suggestion that alphasatellites reduce virus titre, allowing infected plants to survive as sources for onward transmission for a longer period (Briddon et al., 2004; Saunders & Stanley, 1999). More recently, Idris et al. (2011) have shown that a phylogenetically distinct alphasatellite (referred to at the time as DNA 2, but now known as *Ageratum* yellow vein Singapore alphasatellite; Saunders et al., 2002) both attenuates symptoms of a begomovirus–betasatellite infection and preferentially...
Fig. 6. Neighbour-joining phylogenetic dendrogram based upon an alignment of the full-length nucleotide sequences of all full-length alphasatellites available in the databases. The species used were *Ageratum* leaf curl Cameroon alphasatellite (ALCCMA), *Acalypha* yellow vein alphasatellite (AcaYVA), *Ageratum* yellow vein alphasatellite (AYVA), *Ageratum* yellow vein India alphasatellite (AYVINA), *Ageratum* yellow vein Kenya alphasatellite (AYYKEA), *Ageratum* yellow vein Pakistan alphasatellite (AYVPKA), *Ageratum* yellow vein Singapore alphasatellite (AYVSGA), *Euphorbia* mosaic alphasatellite (EuMA), Cotton leaf curl Dabwali alphasatellite (CLCuDaA), Cotton leaf curl Lucknow alphasatellite (CLCuLuA), Cotton leaf curl Multan alphasatellite (CLCuMuA), Cotton leaf curl Shahdadpur alphasatellite (CLCuShA), *Duranta* leaf curl alphasatellite (DuLCA), *Hibiscus* leaf curl alphasatellite (HLCuA), *Malvastrum* yellow mosaic alphasatellite (MalYMA), *Malvastrum* yellow mosaic Cameroon alphasatellite (MalYMCMA), *Malvastrum* yellow mosaic Hainan alphasatellite (MalYVHnA), Melon chlorotic mosaic alphasatellite (MeCMA), Okra leaf curl alphasatellite (OLCuBaA), Okra leaf curl Barombi alphasatellite (OLCuBaA), Okra leaf curl Mali alphasatellite (OLCuMa), Okra yellow crinkle Cameroon alphasatellite (OLCuMa), *Sida* yellow vein alphasatellite (SyYVA), *Sida* yellow vein Vietnam alphasatellite (SyVVNA), Tobacco curly shoot alphasatellite (TbCSA), Tomato leaf curl alphasatellite (ToLCA), Tomato leaf curl Buea alphasatellite (ToLCPK), Tomato leaf curl Cameroon alphasatellite (ToLCMA), Tomato leaf curl Pakistan alphasatellite (ToLCPK), Tomato leaf curl China alphasatellite (TYLCCNA), Tomato leaf curl Yunnan alphasatellite (TYLYnA), *Vernonia* yellow vein alphasatellite (VeYVA), *Gossypium davidsonii* symptomless alphasatellite (GDavSLA), *Gossypium barbifer* symptomless alphasatellite (GBarSLA), *Gossypium mustelinum* symptomless alphasatellite (GMusSLA) and *Verbesina encelioides* leaf curl alphasatellite (VELCuA). The alphasatellite species for which representatives have been isolated from cotton are shown as white text on a red background. The alphasatellites originating from the NW are indicated; those from Africa are underlined. The alphasatellite species for which one isolate (in each case) has been shown to encode a Rep protein with suppressor of PTGS activity (Nawaz-ul-Rehman et al., 2010) are indicated with an asterisk (*). Numbers at significant nodes indicate percentage bootstrap confidence scores (1000 replicates). The significance of the black arrow is discussed in the text.
reduces betasatellite DNA accumulation in plants. It is likely that the amelioration in symptoms is due to the reduced betasatellite titre (thus less βC1 expressed). However, the possible mechanism for preferential reduction of the betasatellite remains unclear, since betasatellite replication depends upon helper begomovirus replication (Saunders et al., 2008). The most surprising recent finding has been the demonstration of suppressor of gene silencing activity for the Rep protein encoded by (at least some) alphasatellites, as will be discussed in the next section.

The CLCuD complex and RNA silencing

RNA silencing (also known as RNA interference) in eukaryotes controls gene expression to regulate development, genome stability, stress-induced responses and defence against molecular parasites such as transposons and viruses (MacLean et al., 2010; Obbard et al., 2009). The importance of RNA silencing in host defence against viruses is evident from the finding that viruses, as a counter defence, have evolved suppressors which act to overcome silencing and all viruses that have been investigated have been shown to encode one or more suppressors (Bivalkarp-Mehla et al., 2011; Voinnet et al., 1999).

Investigation of the interactions of viruses involved in CLCuD with the RNA silencing pathway has lagged behind other geminiviruses. However, recently some major advances have been made. For the most part, investigations with the CLCuD-associated begomoviruses have confirmed earlier results with other viruses. The surprising finding has been that the CLCuD complex may encode four or even five suppressors. Amin et al. (2011a) have shown that the V2, C2, C4 (encoded by CLCuMuV) and the βC1 (encoded by CLCuMuB) have suppressor activity. Additionally the analysis showed that CLCuMuV C4 and CLCuMuB βC1 bind short RNAs, with a preference for the ds and ss forms, respectively, suggesting that these suppressors act to sequester siRNAs and prevent their incorporation into the RNA-induced silencing complex involved in sequence-specific mRNA degradation (Hammond et al., 2000). Although V2 exhibited the strongest suppressor activity, the protein did not appear to act by binding siRNAs (the study was unable, for technical reasons, to investigate the ability of CLCuMuV C2 to bind siRNAs).

Far more surprising was the result from Nawaz-ul-Rehman et al. (2010) which showed that the Rep proteins encoded by two alphasatellites have suppressor activity; the first time this has been shown for a rolling-circle replication–initiator protein. However, these alphasatellites [Gossypium darwinii symptomless alphasatellite (GDarSLA) and Gossypium mustelinum symptomless alphasatellite (GMusSLA)], originated from a very unusual source (exotic Gossypium species maintained over a long period in Multan, Pakistan). Neither of the alphasatellites had been identified prior to this study. Although GDarSLA is typical of the earlier identified alphasatellites (encoding a 315 amino acid Rep protein), GMusSLA is distinct from all previously described alphasatellites and encodes a 263 amino acid Rep protein. It would be premature to conclude from this that all alphasatellites encode Rep proteins with suppressor activity, particularly in view of the fact that Amin et al. (2011a) could identify no suppressor activity for Cotton leaf curl Multan alphasatellite (CLCuMuA; the first alphasatellite identified; Mansoor et al., 1999). The recent identification of one of these unusual alphasatellites (GDarSLA) in plants in Rajasthan (India) suggests that suppressor activity may be a selective advantage for maintenance of (some) alphasatellites by begomovirus-etasatellite complexes (Zaffalon et al., 2012).

Of course the question arises as to why the CLCuD complex appears to require so many suppressors. Amin et al. (2011a) showed that at least two distinct steps of the silencing pathway are targeted by the complex (by C4/βC1 and V2) and these may be required at different times, or in different tissues to effectively circumvent host defence. The betasatellite-encoded βC1 is essential for begomoviruses to symptomatically infect cotton and it is tempting to suggest that it is the suppressor activity of this protein which is the essential function. However, this cannot be the whole story since, even in the presence of CLCuMuB, non-cotton-adapted begomoviruses, such as ToLCNDV and tomato leaf curl virus (neither of which infect cotton naturally or is normally associated with a betasatellite), can transiently infect cotton and induce mild CLCuD symptoms but cannot maintain an infection (Saeed, 2010a, b). Thus not any virus will do and the viruses themselves make a contribution to the infection of cotton (thus are cotton-adapted) even if, for symptoms, the overriding contribution is from the betasatellite (Qazi et al., 2007; Saeed et al., 2008).

Saeed et al. (2007) have shown that CLCuMuB may complement the movement of the bipartite begomovirus ToLCNDV in the absence of the DNA B component and mutagenesis showed that this phenomenon is mediated by βC1. The potexvirus Potato virus X (PVX) p25 protein is one of three ‘triple gene block’ proteins of potexviruses that are required for cell-to-cell movement of the virus (Angell et al., 1996; Beck et al., 1994). It is an RNA helicase (Kalinina et al., 2002) that moves cell-to-cell (Yang et al., 2000), modifies plasmodesmata (Angell et al., 1996) and is a suppressor of RNA silencing (Voinnet et al., 2000). Random mutagenesis of PVX p25 has shown that suppression is required for cell-to-cell movement (Bayne et al., 2005). The authors suggested that the p25 suppressor activity could be required to prevent the spread of a mobile silencing signal, which might prime adjacent cells to counter the incoming virus. Thus it is possible that βC1 could similarly be suppressing a silencing phenomenon that counters virus cell-to-cell spread, rather than acting as a classical virus movement protein. Certainly the ability of βC1 to bind siRNAs (Amin et al., 2011a) and its localization to the cell periphery (Saeed et al., 2007) support this hypothesis, although intercellular movement of βC1 has not so far been demonstrated.

The micro (mi)RNA pathway is the second branch of the RNAi system. miRNAs are endogenous RNAs of ~22 nt
that play important regulatory roles in animals and plants by targeting mRNA for cleavage or translational repression (Bartel, 2004). Several studies have shown that viral suppressors of RNA silencing, including those of begomoviruses, can bind miRNAs (Chapman et al., 2004; Chellappan et al., 2005; Kasschau et al., 2003). However, it is unclear at this time whether virus infection in the miRNA pathway is intentional (thus benefits the virus) or is an unintentional by-product resulting from overlap of the siRNA and miRNA pathways (small RNA-binding suppressor proteins being sequence non-specific). However, it is clear that interfering in the miRNA pathway is a significant mechanism contributing to the symptoms induced by viruses (not just geminiviruses) in plants. An extensive study by Amin et al. (2011b) investigated the effects of the infection of four begomoviruses (including CLCuMuV and CLCuMuB) on the level of 10 miRNAs involved in plant development. Although the four viruses had differing effects on most miRNAs, at least one consistent effect was seen, the moderate to strong upregulation of miR167 by all four viruses. miR167 is complementary to a portion of the sequences of auxin response factor 6 (ARF6) and ARF8 miRNAs. Transgenic Arabidopsis plants overexpressing miR167 were shorter in stature, and showed abnormal reproductive organs and reduced fertility (Ru et al., 2006), features that are very much typical of virus infection. It is thus possible that some of the effects of begomovirus infection on plant development are due to upregulation of miR167.

Amin et al. (2011c) extended their studies with viruses to assess the effects of all genes encoded by four begomoviruses in 10 developmental miRNAs. Although more confusing than the study with viruses, there were some common effects. The most striking was that, in general, the C2/TrAP caused a significant increase in miRNA levels. For two bipartite begomoviruses, transcriptional profiling in Arabidopsis protoplasts has shown that TrAP induces the expression of >100 genes (Trinks et al., 2005). It is thus not inconceivable that C2/TrAP could induce miRNA genes.

**Temporal changes in the Asian CLCuD complex**

During the epidemic of CLCuD in Pakistan in the 1990s, molecular genetic tools were for the first time brought to bear on the problem. Although the involvement of a begomovirus was long suspected, due to the presence of large numbers of B. tabaci on affected plants, experimental insect transmission of the disease between cotton plants was not achieved until 1992 (Briddon & Markham, 2000). The association of a begomovirus with CLCuD in Pakistan was first shown by PCR amplification with primers designed for begomoviruses (Mansoor et al., 1993). The first full-length sequences of CLCuD-associated begomoviruses were produced by Zhou et al. (1998); viruses that we now call CLCuAlV, CLCuKoV and CLCuMuV. In subsequent studies, additional begomovirus species were identified: CLCuRaV, ToLCBaV and PaLCuV (Kirthi et al., 2004; Mansoor et al., 2003b; Zhou et al., 1998). Finally, in 2000, the aetiology of CLCuD in Asia was shown conclusively with Koch’s postulates satisfied for a monopartite begomovirus with a betasatellite, now known as CLCuMuB (Briddon et al., 2001). Additionally, all field-collected plants were shown also to be associated with an alphasatellite (Briddon et al., 2004). Thus the epidemic of CLCuD in Pakistan and north-eastern India during the 1990s involved multiple monopartite begomoviruses, with many plants containing more than one species, a single species of betasatellite and various alphasatellites.

After a period of calm, with no significant losses in the crop due to CLCuD during the late 1990s and early 2000s, it became evident that a change in the virus complex had occurred. All previously resistant cotton varieties started showing symptoms of CLCuD from 2001 onwards (Mansoor et al., 2003a). This signalled the initiation of a second epidemic which rapidly spread to all cotton-growing areas of Pakistan, with the exception of Sindh, and into north-western India. Rapidly, the betasatellite associated with the new epidemic was characterized; the begomovirus took a little longer. The betasatellite was shown to be, for the most part, the CLCuMuB associated with the first epidemic but was recombinant, with a small amount (~100 nt) of sequence of the SCR derived from a betasatellite first identified in tomato: Tomato leaf curl betasatellite. This betasatellite has been referred to as the ‘Burewala strain’ of CLCuMuB [CLCuMuB<sup>Bu</sup>], with the original betasatellite referred to as the Multan strain (CLCuMuB<sup>Mult</sup>); Amin et al., 2006). The underlying basis for the apparent selection of this recombinant betasatellite in resistant cotton varieties remains unresolved. Amin et al. (2006) suggested that the recombinant fragment may enhance interaction with the helper begomovirus. However, since the molecular basis for interaction of the helper begomovirus-encoded Rep with betasatellites remains unclear (as described earlier), this remains a hypothesis.

The begomovirus associated with resistance breaking was also found to be recombinant. This virus species, CLCuBuV, was shown to be a recombinant consisting of sequences derived from CLCuMuV and CLCuKoV. (Amrao et al., 2010b). Surprisingly, CLCuBuV was shown to lack the C2 gene. Although not essential for infectivity (experimentally, viruses with mutations of this gene are infectious to plants; Baliji et al., 2007), the C2 gene (or its homologue the TrAP gene) is found in all dicot-infecting geminiviruses, with the exception of the dicot-infecting mastreviruses for which another gene (RepA) performs most of the C2/TrAP-associated functions; Collin et al., 1996), and no natural mutants lacking the C2/TrAP gene had previously been identified in the field.

**CLCuD and the whitefly vector B. tabaci**

Although, at least for taxonomic purposes, all begomoviruses are transmitted by a single species of whitefly, there
is vector specificity below the B. tabaci species level. For example, Bedford et al. (1994) showed that although the ‘B’ biotype of B. tabaci could transmit all 15 begomoviruses tested, other biotypes could not transmit some begomoviruses and some viruses were transmitted with greater efficiency by some biotypes. Thus, it appears that at least some begomoviruses have evolved to be transmitted by specific biotypes. This may explain why throughout the first CLCuD epidemic in Pakistan, Sindh province remained largely unaffected (Panwar et al., 2001). Central and southern Sindh has only recently begun to see significant incidence of CLCuD in cotton (Mansoor et al., 2006). However, the range of begomoviruses associated with the disease in Sindh is greater than that in the adjacent Punjab province (Amrao et al., 2010a) most probably because, due to the earlier absence of CLCuD, farmers did not see a need to grow resistant varieties. The reasons for the difference between Sindh and the rest of Pakistan, as well as what has changed recently, remain unclear. A possible explanation could be that a different biotype of B. tabaci was recently introduced into Sindh, one that is able to (more) efficiently spread begomoviruses to cotton. Ahmed et al. (2011), using whitefly samples collected across the Punjab and Sindh, showed that two distinct biotypes (described by them as species – see next paragraph) are present in Pakistan, with both present in the Punjab, but only one present in Sindh. However, whether the biotype present in Sindh lacks the ability to transmit CLCuD and whether the other biotype is spreading/has spread into Sindh remains to be determined.

Entomologists have long considered B. tabaci to be a species complex and recently a proposal has been put forward to subdivide B. tabaci into 24 morphologically indistinguishable species, based on molecular sequence data (De Barro et al., 2011). This will undoubtedly lead to a period of uncertainty with respect to taxonomy of begomoviruses, since few, if any, begomoviruses have been assessed for vector specificity at the biotype (proposed species) level. Nevertheless, it will be interesting to investigate the basis for vector specificity with the proposed 24 ‘novel’ species of whitefly.

Countering the CLCuD complex – past, present and future

Prior to 1988 CLCuD was not a significant problem for cotton cultivation. In 1988, something changed and this previously insignificant disease became epidemic. The cultivation of a highly susceptible American cotton variety, S12, is generally accepted to be the major contributor to this change (Briddon & Markham, 2000). This implies that the earlier cultivated (local) varieties had some level of resistance to the virus(es) causing CLCuD and that, upon infection of a variety with no resistance, the virus complex was able to adapt for infection of and spread in cotton. Although the virus causing CLCuD during the early days of this first epidemic was not characterized, the widespread occurrence of CLCuMuV, and its contribution to many of the begomovirus species identified in cotton in Asia (by recombination), suggests that this species might be the original virus that infected cotton in conjunction with CLCuMuB.

Conventional selection/breeding approaches during the 1990s yielded varieties with excellent resistance to the ‘Multan’ strain of CLCuD. Near universal cultivation of these resistant varieties across Pakistan (with the exception of Sindh) returned CLCuD to obscurity. A study of the donors of the resistance, varieties LRA5166 and CP-15/2, has suggested that three genes are involved, two for resistance and a further gene imparting suppression of the resistance (Rahman et al., 2005). However, the nature of these resistance genes and their precise mechanism of action remain unresolved.

Despite almost 10 years of effort since the appearance of the ‘Burewala’ strain of CLCuD, conventional breeding has not been entirely successful in yielding resistance to the disease (Farooq et al., 2011). Some promising lines with good tolerance, but that still support virus infection, exhibit mild symptoms and nevertheless provide acceptable yields, have been identified (Rahman & Zafar, 2007). However, the desired immune lines have not been forthcoming. Tolerant lines, although better than no resistance at all, are a problem since they still support virus replication and systemic spread. They thus can act as reservoirs for onwards transmission of the virus and virus infection may select for more aggressive (destructive) variants.

In common with many cultivated plant species (such as tomato, chili peppers, maize, cassava and potato), G. hirsutum has its origins in the NW and is believed to have been introduced to the OW during the 1800s. Its cultivation was encouraged in the OW colonies during the American Civil War when demand, particularly from England, could not be met from the NW. This led to the near total replacement of G. arboreum (the native cotton species which is immune to CLCuD) by G. hirsutum (which is both higher yielding and produces a higher quality of cotton) on the Indian subcontinent. Almost universally, the introduced NW crops have come to suffer extensively from geminiviruses. The most prominent examples here are cassava across Africa and parts of India being affected by cassava mosaic begomoviruses (Patil & Fauquet, 2009) and maize being affected by Maize streak virus across Africa (Shepherd et al., 2010). This is usually attributed to the fact that these crops are not affected by geminiviruses in the NW (no geminiviruses are known to infect either cassava or maize in the NW) or that the viruses that affect them in the NW are genetically distinct from those in the OW. These crops would thus not have had a long evolutionary association with the geminivirus pathogens in the OW and would be unlikely to have evolved adequate defences (resistance) against them.

One possible answer to the lack of genetic diversity (with respect to pathogen resistance) in many crop species is genetic
engineering using either the pathogen-derived approach (with the engineered resistance being sequence-derived from the pathogen itself) or with the sequence derived from other, heterologous sources (the non-pathogen-derived approach). For geminiviruses, numerous pathogen-derived and non-pathogen-derived approaches to achieving resistance in planta have been investigated (reviewed by Ilyas et al., 2011; Vanderschuren et al., 2007). Specifically, for viruses associated with CLCuD, studies have investigated RNA silencing-mediated resistance. Asad et al. (2003) investigated expression of a truncated Rep gene of CLCuKoV in antisense orientation, in the model plant Nicotiana tabacum, and found that this protects plants from virus infection in some lines. Similar constructs were transformed into cotton and shown to delay and/or ameliorate virus symptoms following exposure to viruliferous insects (Hashmi et al., 2011).

In addition to breeding for disease tolerance against CLCuD in cotton, other conventional approaches, including controlling the vector whitefly, eradication of potential alternative weed hosts, seed treatment and various agronomic approaches have proven useful in reducing losses (Farooq et al., 2011).

**What does the future hold?**

It is evident that the globalization of agriculture will increasingly lead to the spread of viral diseases. There are numerous examples, just for geminiviruses, where this has already occurred. *Tomato yellow leaf curl virus*, a monopartite begomovirus with its origins in the Mediterranean/Middle East, has spread across southern Europe, North Africa and to the NW, Asia and Australia (Lefeuvre et al., 2010); *Squash leaf curl virus*, a begomovirus with its origins in North America, has been introduced into the Middle East (Idris et al., 2006); and *Bean yellow dwarf virus*, a mastrevirus with its origins in the Middle East/Southern Asia, has spread to southern Africa, most probably by the migration of people from southern Asia into Africa (Nahid et al., 2008). It would thus be naïve to assume that similar things could not happen with CLCuD. All areas where cotton is grown have environmental conditions that are suitable for the proliferation of the disease agents and host populations of B. tabaci, the vector of begomoviruses.

Recent evidence suggests that we are already seeing the first instances of the spread of CLCuD. The African CLCuD-associated begomovirus, CLCuGeV, has been identified in cotton in southern Pakistan (Tahir et al., 2011). Although not a significant problem at this time, any increase in virus diversity, particularly the introduction of a genetically distinct virus (thus the introduction of ‘new blood’), is worrying in an area that already is affected by a large diversity of begomoviruses. The only saving grace is that, at least up to now, the African cotton betasatellite, Cotton leaf...
curl Gezira betasatellite, has not been found in Pakistan and there is no evidence for recombination between the African and Asian viruses.

Far more worrying is the recent outbreak of CLCuD in southern China (Cai et al., 2010), geographically far removed from the areas of Pakistan and India that are affected. Significantly, this area of China was not previously home to geminiviruses (at least none were reported) and was not until recently a cotton-growing area. The disease was first reported in Hibiscus rosa-sinensis (Mao et al., 2008; Fig. 7) and subsequently okra (Xie et al., 2012). In each case, the disease was shown to involve CLCuMuV and CLCuMuB<sup>Mal</sup>. Analysis of the sequences showed them to be most closely related to virus and betasatellite isolates identified in India/Pakistan in <i>H. rosa-sinensis</i> (Figs 3, 4 and 6). Three strains of CLCuMuV are recognized at this time and the isolates in China are all of the ‘Faisalabad’ strain – the strain prevalent in Pakistan prior to resistance breaking (Fig. 3). <i>H. rosa-sinensis</i> is a well known alternative host of the CLCuD complex and this strongly suggests that the disease was introduced into China in infected <i>H. rosa-sinensis</i> from India or Pakistan. Pakistan exports ornamental plants, including <i>H. rosa-sinensis</i>, mostly to the Gulf States, and the source plants from which cuttings are taken are almost universally symptomatic (R.W. Briddon, personal observation). The sequence evidence for the CLCuD outbreak in China indicates that it is the ‘Multan’ strain (Fig. 3) that has been introduced. This suggests that the resistance seen in Pakistan during the 1990s (Rahman et al., 2005) could potentially be useful in controlling the outbreak, at least in cotton.

If previous experience teaches us anything, it is that resistance to the viruses causing CLCuD based upon a single mechanism of action (be that conventional host-plant resistance or non-conventional, transgenic approaches) are unlikely to be durable; the resistance introduced in Pakistan and India during the late 1990s lasted less than 5 years. The only way to overcome this problem will be to ‘stack’ multiple resistances, based upon distinct mechanisms of action (Ilyas et al., 2011). Thus, for example, reinforce the best natural resistance with an RNAi-based transgenic resistance construct and/or one based on the non-pathogen-derived approach. This stacking of resistance types is far less likely to be overcome than any resistance based on a single mechanism of action. Also, there need to be far stricter controls over the movement of agricultural products and live ornamental plants that can potentially harbour the viruses and satellites that cause CLCuD, as well as the whitefly that transmits them, to avoid a recurrence of the situation in southern China. Farmers and extension staff in all cotton-growing regions need to be made aware of CLCuD to ensure early identification and the timely implementation of control measures, should the disease be introduced. A good example here is the information distributed by Plant Health Australia (Gambley, 2010).

**Acknowledgements**

R.W.B. is supported by the Higher Education Commission (HEC, Govt of Pakistan) under the ‘Foreign Faculty’ scheme. M.N.S. is supported by the HEC/Swedisth Institute under the ‘Overseas Scholarship Scheme for PhD in Selected Fields’. The authors thank Professor Yule Liu for providing photographs of plants affected by CLCuD in China.

**References**


against systemic infection by plant viruses with a triple gene block. Proc Natl Acad Sci USA 91, 10310–10314.


Cotton leaf curl disease


