

Article

## Regional Changes in the Sequence of Cotton Leaf Curl Multan Betasatellite

Sohail Akhtar <sup>†</sup>, Muhammad Nouman Tahir, Ghulam Rasool Baloch, Shaista Javaid, Ali Qaiser Khan, Imran Amin, Rob W. Briddon and Shahid Mansoor <sup>\*</sup>

Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan; E-Mails: sakhtar.pp@uaf.edu.pk (S.A.); nouman.nibge@yahoo.com (M.N.T.); grasoolbaloch@yahoo.com (G.R.B.); shaistajavid85@gmail.com (S.J.); kk\_684@yahoo.com (A.Q.K.); imranamin1@yahoo.com (I.A.); rob.bridon@gmail.com (R.W.B.)

<sup>†</sup> Present address: Sub-campus Burewala, University of Agriculture, Faisalabad, Pakistan.

<sup>\*</sup> Author to whom correspondence should be addressed; E-Mail: shahidmansoor7@gmail.com; Tel.: +92-41-9201804; Fax: +92-41-9201801.

Received: 24 March 2014; in revised form: 2 May 2014 / Accepted: 12 May 2014 /

Published: 23 May 2014

---

**Abstract:** Cotton leaf curl disease (CLCuD) in Pakistan and northwestern India is caused by monopartite begomoviruses in association with an essential, disease-specific satellite, Cotton leaf curl Multan betasatellite (CLCuMB). Following a recent upsurge in CLCuD problems in Sindh province (southern Pakistan), sequences of clones of CLCuMB were obtained from Sindh and Punjab province (central Pakistan), where CLCuD has been a problem since the mid-1980s. The sequences were compared to all sequences of CLCuMB available in the databases. Analysis of the sequences shows extensive sequence variation in CLCuMB, most likely resulting from recombination. The range of sequence variants differ between Sindh, the Punjab and northwestern India. The possible significance of the findings with respect to movement of the CLCuD between the three regions is discussed. Additionally, the lack of sequence variation within the only coding sequence of CLCuMB suggests that the betasatellite is not involved in resistance breaking which became a problem after 2001 in the Punjab and subsequently also in northwestern India.

**Keywords:** begomovirus; betasatellite; genetic diversity; recombination

---

## 1. Introduction

Cotton leaf curl disease (CLCuD) is a major constraint for cotton production across central Pakistan and northwestern areas of India [1]. The disease appeared in an epidemic form in the early 1990s in the vicinity of the city of Multan, Pakistan [2]. In the late 1990's the introduction of resistant cotton varieties developed by conventional breeding methods restored Pakistani cotton production to pre-epidemic levels. Unfortunately, from 2001 onwards, the disease reappeared on all previously resistant varieties in the vicinity of the city of Burewala and rapidly spread across most cotton growing areas of Pakistan and into northwestern India [3,4]. CLCuD is caused by begomoviruses in association with a disease-specific satellite, Cotton leaf curl Multan betasatellite (CLCuMB) [5,6].

Viruses of the genus *Begomovirus* (family *Geminiviridae*) are transmitted by the whitefly *Bemisia tabaci*, with genomes consisting of either one or two small (~2.8 kb), circular single-stranded (ss)DNA components encapsidated in twinned, quasi-isometric particles [7]. In the Old World (OW) a small number of bipartite begomoviruses have been identified, the majority instead being monopartite. This contrasts with the New World where begomoviruses with bipartite genomes are the norm; the first native monopartite begomovirus only having been identified recently [8,9]. In the OW the majority of monopartite are associated with a class of ssDNA satellites collectively known as betasatellites [10,11].

Betasatellites are small (~1.4 kb), circular, ssDNA satellites that have only been identified in the OW, are associated with monopartite begomoviruses, and depend upon a helper virus for replication, movement in and transmission between plants [6,12,13]. The sequences of betasatellites have three major features—a single conserved (in both position and sequence) gene (referred to as  $\beta$ C1), a region of sequence rich in adenine (A-rich) and a sequence conserved among all betasatellites (known as the satellite conserved region (SCR) [12,14]. The SCR contains a predicted stem-loop structure containing the nonanucleotide sequence TAATATTAC which, for geminiviruses, is the origin of virion-strand DNA replication [15]. The  $\beta$ C1 gene typically encodes a protein of 118 amino acids that mediates all betasatellite functions.  $\beta$ C1 is a pathogenicity determinant, a suppressor of post-transcriptional gene silencing (PTGS; a small RNA-mediated host defense), binds DNA and possibly mediates virus movement in plants [16–21]. The function of the A-rich region is unclear. It has been suggested that this may be a “stuffer” to increase the size of the molecule for encapsidation [12,22].

CLCuD across central Pakistan and northwestern India during the 1990s was associated with at least six distinct monopartite begomoviruses but only a single betasatellite, referred to as the “Multan” strain of CLCuMB (CLCuMB<sup>Mul</sup>) [6]. Following resistance breaking, the disease across central Pakistan was found to be associated with only a single virus, a recombinant species, *Cotton leaf curl Burewala virus* (CLCuBuV), consisting of sequences derived from two virus species present in cotton before resistance breaking [23]. The betasatellite associated with CLCuBuV was also recombinant, with a small replacement of sequences within the SCR derived from a distinct betasatellite, Tomato leaf curl betasatellite (ToLCB) [23,24]. This is now referred to as the “Burewala” strain of CLCuMB (CLCuMB<sup>Bur</sup>).

Until recently CLCuD was only a minor, sporadic problem across central and southern Sindh province, Pakistan. For this reason CLCuD resistant cotton varieties were not widely grown—farmers having to make a choice between the high yielding, but susceptible varieties, or playing safe with the lower yielding resistant varieties. In 2005 there was an upsurge of CLCuD in Sindh which was shown

to be associated with a virus species present in cotton in the Punjab pre-resistance breaking, a new recombinant virus distinct from CLCuBuV, CLCuMB<sup>Bur</sup> and a new strain of CLCuMB containing a smaller recombinant fragment from ToLCB, now referred to as the Shahdadpur strain of CLCuMB (CLCuMB<sup>Sha</sup>) [25]. In northwestern India two recent studies have shown CLCuBuV to dominate in cotton post-resistance breaking but a virus common prior to resistance breaking, *Cotton leaf curl Rajasthan virus* (CLCuRaV) was also detected infrequently [4,26]. Although both studies showed the presence of CLCuMB, neither determined the strain thereof.

The study presented here has analyzed the sequences of CLCuMB recently isolated from Sindh and the Punjab (Pakistan), for comparison to isolates obtained earlier, and has identified region-specific sequence changes. The significance of these findings is discussed.

## 2. Materials and Methods

### 2.1. Sample Collection

Leaves of cotton plants showing symptoms typical of CLCuD were collected from fields around the Central Cotton Research Institute (CCRI) Multan (Punjab, Pakistan) in 2008/2009 and areas of Sindh in 2010 sampled in previous studies [25,27].

### 2.2. PCR-Mediated Amplification and Cloning of Betasatellites

Total genomic DNA was extracted from leaves using the CTAB method [28] and stored at  $-20\text{ }^{\circ}\text{C}$ . Betasatellites were amplified by PCR using universal primers [29]. PCR products were cloned in pTZ57R/T (Fermentas) and sequenced commercially, in both orientations, using the primer walking strategy (Macrogen Inc., Seoul, Korea).

### 2.3. Sequence Analysis

Sequences were assembled and analysed using the Lasergene (DNASTAR Inc., Madison, Wisconsin, USA) sequence analysis package. Sequence similarity searches (BLAST) were performed online by comparing the sequences characterized in this study with sequences available in the databases using BLAST. Multiple sequence alignments were performed using MegAlign (Lasergene) and ClustalX2 [30]. Phylogenetic dendrograms were constructed using the Neighbour-joining algorithm of ClustalX2 and viewed, manipulated and printed using Treeview [31]. Potentially recombinant sequences were identified using the Recombination Detection Program version 3 (RDP version 3; [32]).

## 3. Results

### 3.1. Analysis of Betasatellite Sequences

A total of 11 presumed full-length (~1.4 kb; 4 from the Punjab and 7 from Sindh) and 26 defective (~0.7 kb; 2 from the Punjab and 24 from Sindh) betasatellites were cloned and sequenced. The sequences are available in the nucleotide sequence databases under the accession numbers given in Tables 1 and 2. The complete nucleotide sequences of betasatellites were compared to sequences of

betasatellites available in the database using BLAST. The comparison revealed sequence identities ranging from 96% to 99% with available isolates of CLCuMB. Since the species demarcation threshold for betasatellite is 78% [33], this indicates that all betasatellites cloned from cotton here are isolates of CLCuMB.

**Table 1.** Origins and features of full-length betasatellites.

Isolate descriptor	Origin	Accession number	Size (nt)	Position of $\beta$ C1 gene (coordinates of start/stop codons)
[PK:Hala:09]	Hala/Sindh	HE601944	1350	195–551
[PK:Hala1:09]	Hala/Sindh	HE601945	1350	195–551
[PK:Sha:06]	Shahdadpur/Sindh	HE601941	1375	195–551
[PK:Tjam2:09]	Tandojam/Sindh	HE601946	1364	195–551
[PK:Tjam1:09]	Tandojam/Sindh	HE601947	1348	195–551
[PK:Tjam:09]	Tandojam/Sindh	HE601948	1349	194–550
[PK:Tjam:06]	Tandojam/Sindh	HE601940	1350	195–551
[PK:Mul:08]	Multan/Punjab	HE601938	1345	195–551
[PK:Mul1:08]	Multan/Punjab	HE601939	1349	195–551
[PK:Mul1:09]	Multan/Punjab	HE601942	1350	195–551
[PK:Mul2:09]	Multan/Punjab	HE601943	1355	195–551

A phylogenetic analysis, based upon alignment of the full-length sequences determined here with available CLCuMB sequences in the databases is shown in Figure 1. The dendrogram shows the satellites to segregate into two major groups. The first group contains the non-recombinant CLCuMB associated with the CLCuD epidemic of the 1990s (CLCuMB<sup>Mul</sup>) first identified by Briddon *et al.* [5], as well as the recombinant CLCuMB recently identified in the Sindh province (Pakistan; CLCuMB<sup>Sha</sup>, [25]). Parts of group 1 are also some CLCuMB isolates, labeled subgroup 3 (SG3) in Figure 1, which group with the CLCuMB<sup>Mul</sup>/CLCuMB<sup>Sha</sup> isolates but are basal to them. The second group contains the CLCuMB sequences first identified with resistance breaking in cotton identified by Amin *et al.* [24], known as the Burewala strain (CLCuMB<sup>Bur</sup>). The CLCuMB isolates characterized as part of the study here fall into both major groups, although none segregate with the non-recombinant CLCuMB<sup>Mul</sup>.

### 3.2. Variation in SCR Sequences

An alignment of part of the SCR sequences of the betasatellite isolates obtained here (including the defective clones) with selected sequences from the databases is shown in Figure 2. The alignment shows the recombinant fragment (~105 nt; grey bar below the diagram), believed to have originated from ToLCB that distinguishes CLCuMB<sup>Bur</sup> isolates from CLCuMB<sup>Mul</sup> isolates [23,24]. An RDP analysis (Supplementary Figure S1, Supplementary Table S1) actually suggests that, rather than just spanning the region indicated in Figure 2, the recombination encompasses also the conserved nonnucleotide-containing hairpin for SG1 and SG2A isolates, whereas for SG2B isolates the recombinant fragment is ~150 nt spanning the area indicated in Figure 2. For some of the isolates identified by Amrao *et al.* [25], and also for some isolates identified here, all originating from Sindh,

the recombinant fragment of ToLCB is somewhat smaller (~25 nt; grey bar labeled SG4) and was previously designated the “Shahdadpur” strain of CLCuMB (CLCuMB<sup>Sha</sup>). The alignment shows an additional group of CLCuMB isolates that contain a ~27 nt fragment derived from ToLCB that overlaps with the fragment in the SG4A isolates (labeled as SG4B and SG3 in Figure 2). These isolates differ in their origins, the SG4B originating from Sindh and SG3 isolates from the Punjab. The presence of these smaller recombinant fragments of ToLCB is supported by the RDP analysis but only by one of the detection methods (GENECONV; Supplementary Table S1).

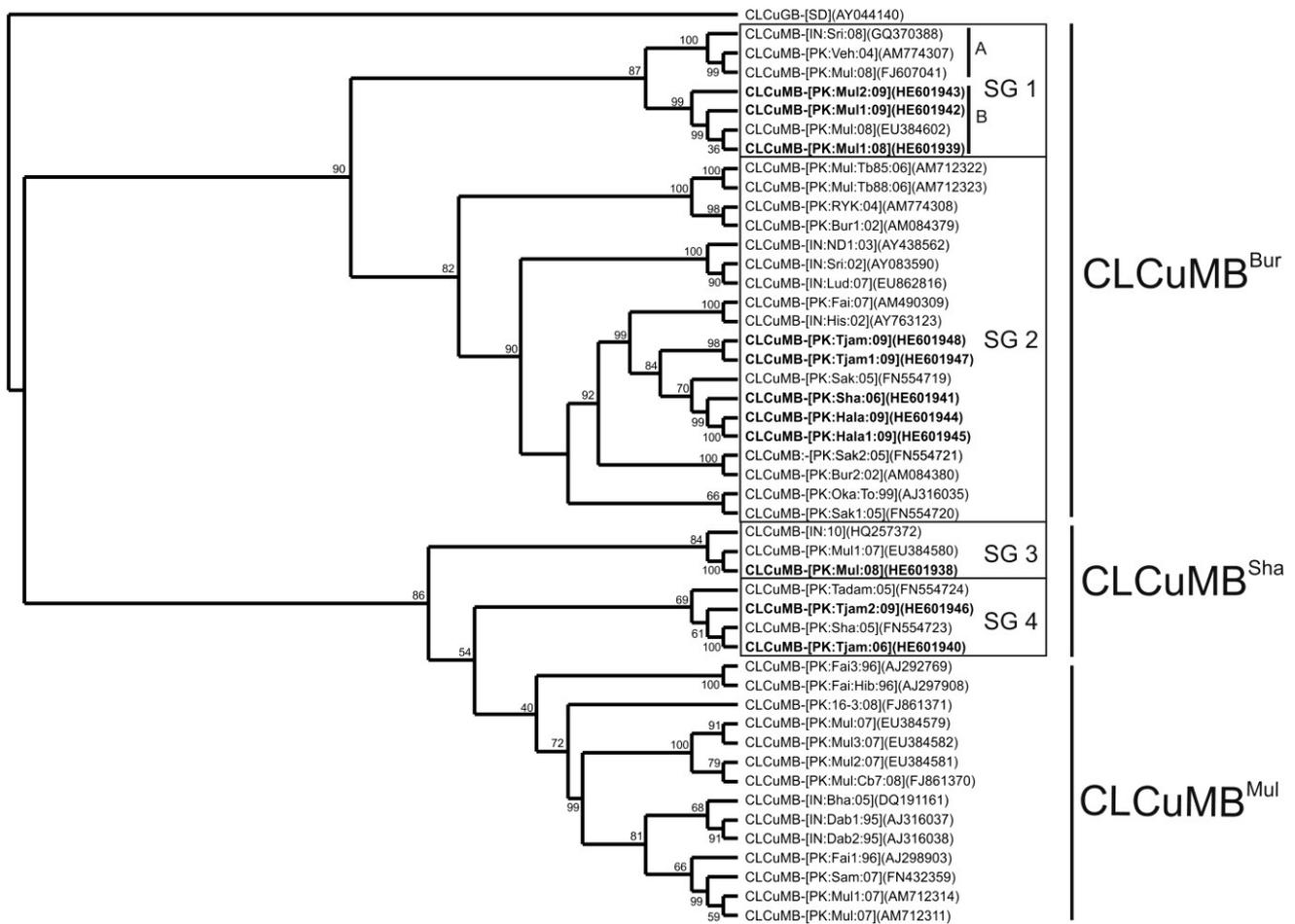
**Table 2.** Origins and features of defective betasatellites.

Isolate descriptor	Origin	Accession number	Size (nt)	Coordinates of deleted region *
[PK:Sak:06]	Sakrand/Sindh	HE602942	645	048–758
[PK:Sak1:06]	Sakrand/Sindh	HE602943	690	065–731
[PK:Sak2:06]	Sakrand/Sindh	HE602944	668	065–735
[PK:Sha1:06]	Shahdadpur/Sindh	HE602945	727	096–721
[PK:Hala:09]	Hala/Sindh	HE602950	788	202–763
[PK:Hala1:09]	Hala/Sindh	HE602951	788	202–763
[PK:Hala3:09]	Hala/Sindh	HE602953	818	208–740
[PK:Hala4:09]	Hala/Sindh	HE602956	699	110–763
[PK:Mat1:09]	Matiari/Sindh	HE602957	681	099–786
[PK:Hala5:09]	Hala/Sindh	HE602958	694	097–753
[PK:Tjam:09]	Tandojam/Sindh	HE602959	688	116–780
[PK:Sha:06]	Shahdadpur/Sindh	HE602934	680	157–832
[PK:Sha2:06]	Shahdadpur/Sindh	HE602948	678	130–801
[PK:Hala2:09]	Hala/Sindh	HE602952	712	053–696
[PK:Tjam:06]	Tandojam/Sindh	HE602935	678	129–805
[PK:Tjam1:06]	Tandojam/Sindh	HE602936	678	129–805
[PK:Tadam:06]	Tandoadam/Sindh	HE602937	713	129–769
[PK:Tadam1:06]	Tandoadam/Sindh	HE602938	697	115–771
[PK:TAyar:06]	TandoAllahyar/Sindh	HE602939	680	116–780
[PK:TAyar1:06]	TandoAllahyar/Sindh	HE602940	735	153–773
[PK:TAyar2:06]	TandoAllahyar/Sindh	HE602941	690	084–748
[PK:TAyar3:06]	TandoAllahyar/Sindh	HE602946	685	117–805
[PK:TAyar4:06]	TandoAllahyar/Sindh	HE602947	809	151–714
[PK:Mat:09]	Matiari/Sindh	HE602949	749	115–732
[PK:Mul:09]	Multan/Punjab	HE602954	703	105–754
[PK:Mul1:09]	Multan/Punjab	HE602955	953	224–631

\* Relative to isolate CLCuMB-[PK:Sak:05] (FN554719).

The defective (deletion mutant) CLCuMB clones are all approx. half the size (630 to 700 bp) of full-length betasatellites (typically ~1350 bp; [11,12]). The sequences deleted cover the  $\beta$ C1 gene and its promoter. In each the SCR, the A-rich and the sequences between these (across the conserved hairpin structure) are maintained (Table 2).

**Figure 1.** Phylogenetic analyses of Cotton leaf curl Multan betasatellite (CLCuMB) sequences. Shown is a neighbor-joining phylogenetic dendrogram based upon alignment of the complete nucleotide sequences of CLCuMB isolates produced as part of the study presented here (highlighted with bold text) with selected CLCuMB sequences available in the databases. The alignment was arbitrarily rooted on the sequence of Cotton leaf curl Gezira betasatellite (CLCuGB-[SD]), a distantly related betasatellite originating from Sudan, as an out group. In each case, the isolate descriptor and accession number is given. Numbers at the nodes are percentage bootstrap confidence values (1000 replicates). The recombinant CLCuMB<sup>Bur</sup>, the recombinant CLCuMB<sup>Sha</sup> and non-recombinant CLCuMB<sup>Mul</sup> isolates are indicated. The subgroups (SG) are discussed in the text.



Not considered in the previous analyses of the SCR sequences of CLCuMB was the sequence immediately upstream of the ToLCB insertions. For the recently characterized isolates, including those identified here, these sequences differ from those of CLCuMB<sup>Mul</sup>. For many of the CLCuMB<sup>Bur</sup> isolates from Pakistan in the earlier studies [23,24] these sequences are homologous to the those of CLCuMB<sup>Mul</sup> (the SG1 isolates) characterized earlier. The sequence replacements in this region not homologous to CLCuMB<sup>Mul</sup> are of two types (shown as either solid black lines or dashed black lines in Figure 2). The origins of the sequences indicated by dashed lines are unclear, having some levels of sequence identity to a number of distinct betasatellite species and could result from recombination, although the RDP analysis did not support this possibility. Overall the isolates segregate, with respect

to these replacements, according to the phylogenetic groupings identified in Figure 1. Interestingly the SG2A isolates are exclusively from either Sindh or India whereas the SG2B isolates, with a smaller sequence insertion, come only from the Punjab.

An alignment of sequences downstream of the hairpin, falls (at least in part) within the SCR [12] of the betasatellite isolates obtained here (including the defective clones) with selected sequences from the databases shown in Figure 3. The alignment shows the CLCuMB<sup>Bur</sup> SG1 and CLCuMB<sup>Sha</sup> SG3 isolates share a sequence of *ca.* 123 nt which is distinct from the other isolates. The 123 nt sequence shows no significant levels of sequence identity to other sequences in the databases, is not highlighted by the RDP analysis and its origin thus remains unclear. The isolates with this sequence originate from both regions of Pakistan and northwestern India.

Overall, of 147 recombinant CLCuMB isolates available in the databases, the dominant form in India is SG2A (8 SG1A, 3 SG1B, 32 SG2A and 2 SG4A) whereas SG2B and SG3 dominate in the Punjab (4 SG1A, 15 SG1B, 5 SG2A, 17 SG2B and 18 SG3) and Sindh has only SG2A (31) and SG4 (9 SG4A and 3 SG4B). Sindh, despite the low numbers of CLCuMB sequences available, has variants in common with the Punjab (SG2A), but not the variant common in the Punjab (SG2B), and variants in common with India (SG2A and SG4; SG4 has not been identified in the Punjab).

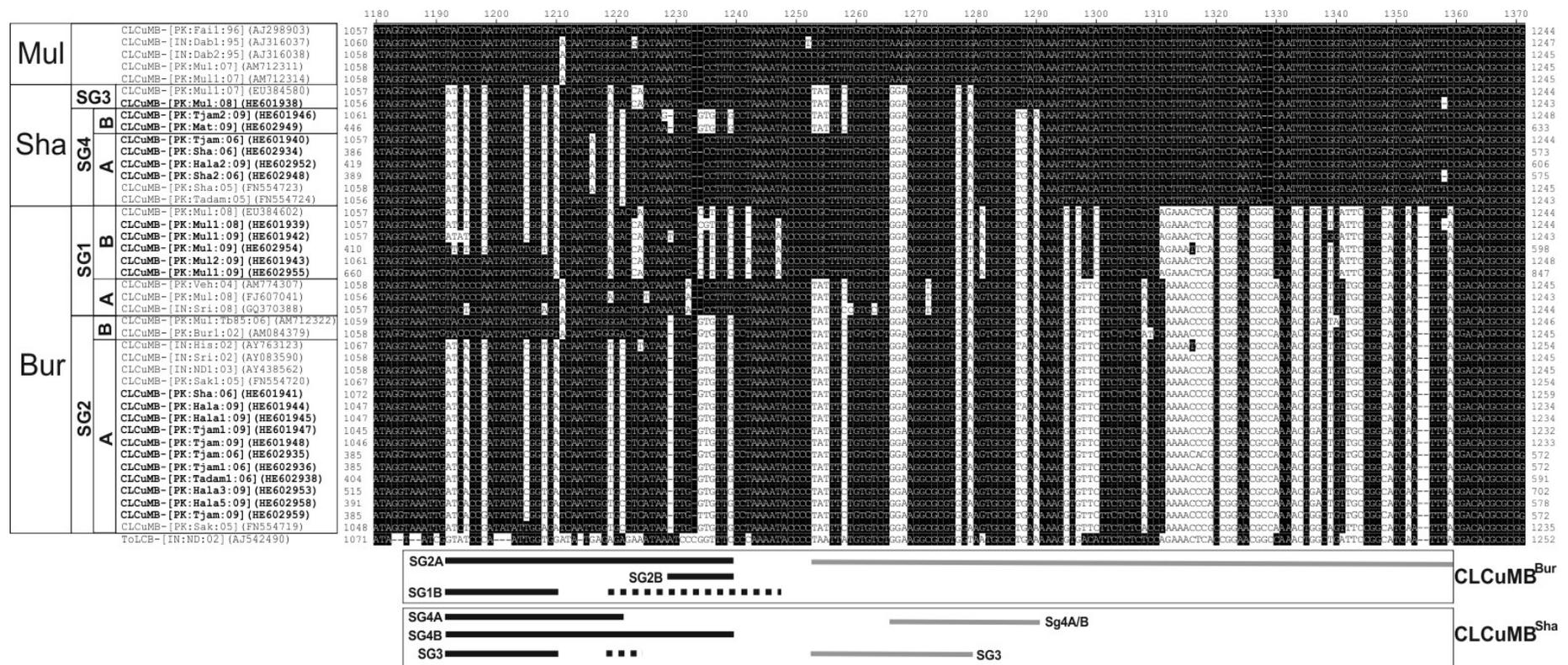
### 3.3. Analysis of CLCuMB Sequences between the A-Rich Region and the SCR

An alignment of CLCuMB (including the defective clones) sequences between the SCR and A-rich region is shown in Figure 4. This shows two clones (CLCuMB-[PK:Fai:07] AM490309 and CLCuMB-[IN:His:02] AY763123) to have a large insertion/deletion (indel) with respect to the other isolates (coordinates 1000 to 1177 of the alignment in Figure 4). The origin of the insertion is unclear since there is no significant sequence identity to sequences in the database except of the CLCuMB isolates shown here. With the exception of these two clones only CLCuMB isolates originating from Sindh have the indel, although the lengths thereof differ, with the sequence being progressively truncated. Although CLCuMB-[IN:His:02]AY763123 was isolated from cotton in India, CLCuMB-[PK:Fai:07] AM490309 was isolated from tomato and shown to be associated with *Cotton leaf curl Rajasthan virus* [34], a begomovirus shown to be extensively associated with CLCuD in India [4] but not identified in *G. hirsutum* in Pakistan.

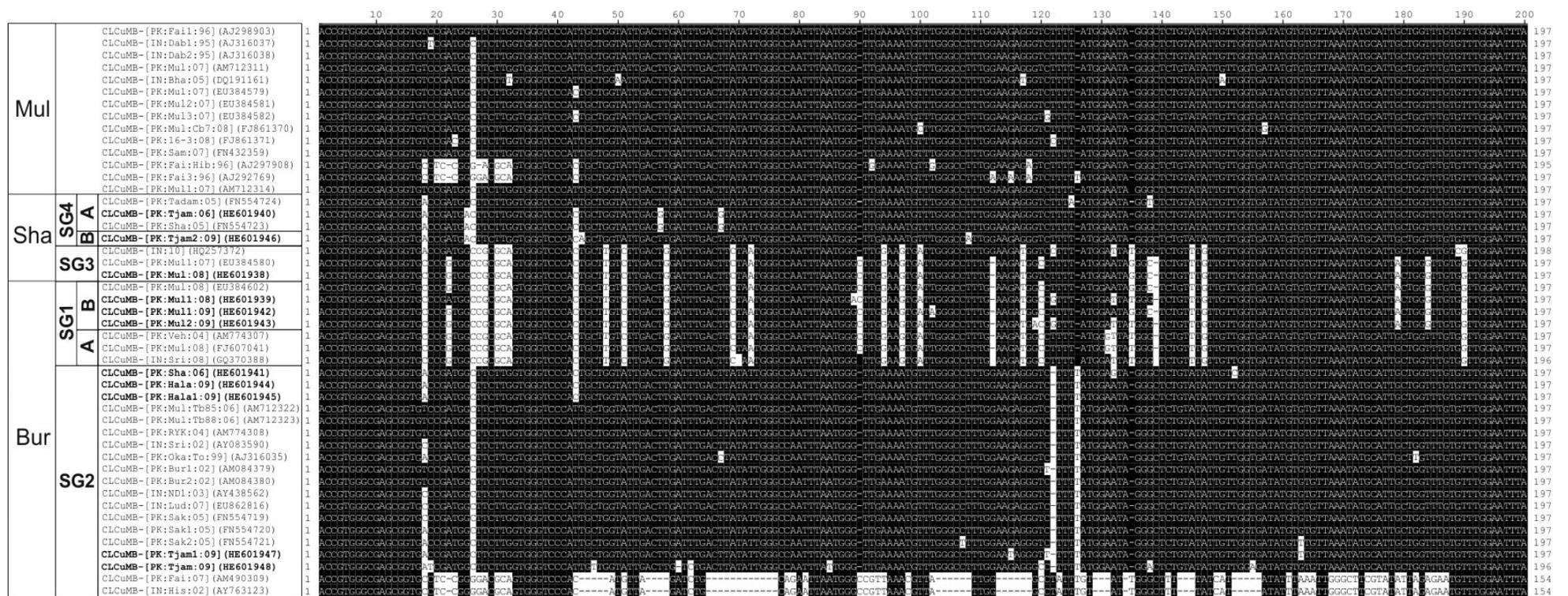
### 3.4. Sequence Variation in the Predicted Amino Acid Sequences of $\beta$ C1

An alignment of the predicted amino acid sequences of the  $\beta$ C1 protein of selected CLCuMB isolates is shown in Figure 5, which shows there is little sequence variation in the protein. Most of the major conserved (across multiple isolates and thus likely to be meaningful rather than random sequence variation or sequencing errors) sequence differences are conservative with respect to amino acid properties. Nevertheless, for the majority of isolates there are amino acid changes that differ between isolates originating from the Punjab and Sindh.

**Figure 2.** Alignment of the nucleotide sequences of the satellite conserved region (SCR) of CLCuMB isolates derived from the study presented here (highlighted with bold text) with selected sequences from the database. The defective isolates were included in this analysis. The homologous sequence of a Tomato leaf curl betasatellite (ToLCB) isolate was included for comparison. Gaps (-) were introduced into sequences to optimize the alignment. In each case, the isolate descriptor and accession number is given. Sequences differing from CLCuMB-[PK:Fail:96] (AJ298903; a CLCuMB<sup>Mul</sup> isolate) are highlighted as black text on a white background. The strain of CLCuMB (Burewala [Bur], Multan [Mul] or Shahdampur [Sha]) is shown on the left as well as the subgroups (SG) identified in Figure 1. The origins of sequences are shown in the two boxes below the alignment for the recombinant CLCuMB isolates (CLCuMB<sup>Bur</sup> and CLCuMB<sup>Sha</sup>). Sequences likely originating from ToLCB are shown with grey bars. Sequences homologous to those of CLCuMB<sup>Bur</sup> are indicated by solid black bars. Sequences of unknown origin are shown by dashed black lines. All other sequences are homologous to CLCuMB<sup>Mul</sup>.

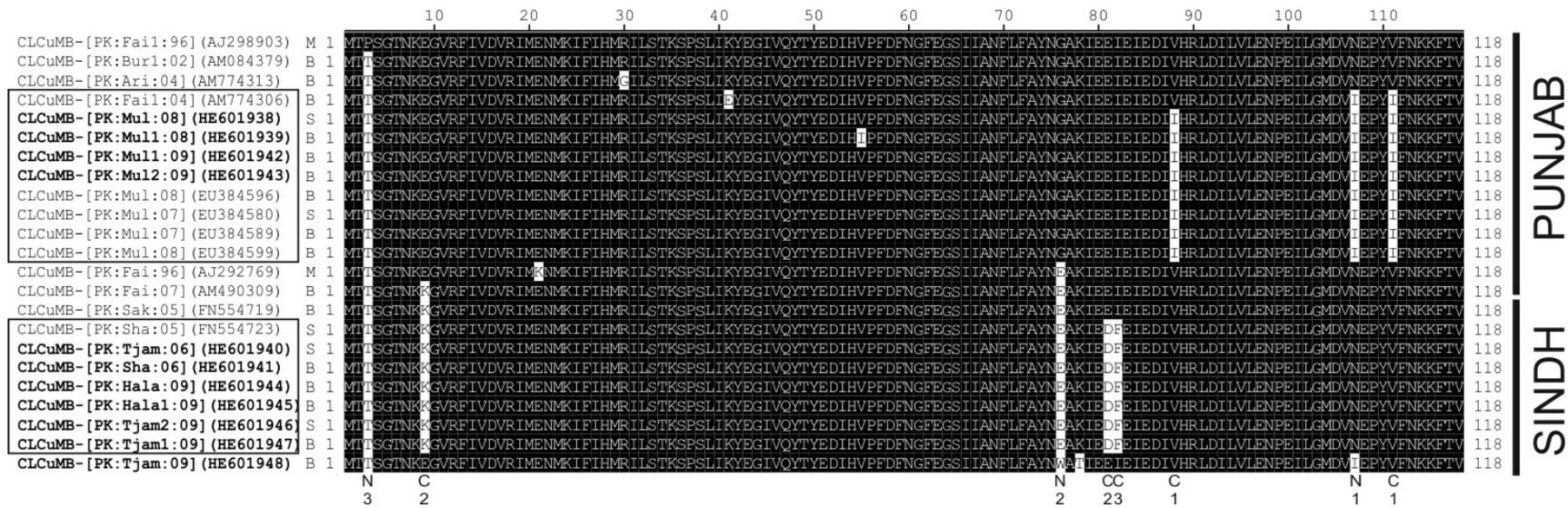


**Figure 3.** Alignment of the nucleotide sequences immediately downstream of the hairpin structure of CLCuMB isolates derived from the study presented here (highlighted with bold text) with selected sequences from the database. The sequence falls, at least in part, within the SCR. Gaps (-) were introduced into sequences to optimize the alignment. In each case, the isolate descriptor and accession number is given. Sequences differing from CLCuMB-[PK:Fail:96] (AJ298903; a CLCuMB<sup>Mul</sup> isolate) are highlighted as black text on a white background. The strain of CLCuMB (Burewala [Bur], Multan [Mul] or Shahdadpur [Sha]) is shown on the left as well as the subgroups (SG) identified in Figure 1.





**Figure 5.** Alignment of the predicted amino acid sequences of the  $\beta$ C1 proteins of selected CLCuMBs. Sequences with the same, or similar, sequence changes have been grouped. In each case, the isolate descriptor, accession number and type (CLCuMB<sup>Bur</sup> [B], CLCuMB<sup>Mul</sup> [M], or CLCuMB<sup>Sha</sup> [S]) are given. Betasatellites produced as part of the study presented here are marked in bold. Sequences are indicated as originating from either the Punjab or Sindh on the right. Standard IUPAC/IUBMB single letter amino acid codes are used. Along the bottom the major amino acid sequence changes are indicated as either conservative (C) or non-conservative (N) with respect to amino acid properties (C1-I and V non-polar/hydrophobic; C2-D,K and E polar/hydrophilic; F and I-non-polar/hydrophobic; N1-I non-polar/hydrophobic whereas N is polar/hydrophilic; N2-G and W non-polar/hydrophobic whereas E is polar/hydrophilic; N3-T polar/hydrophilic whereas P is non-polar/hydrophobic).



#### 4. Discussion

CLCuD continues to adversely affect cotton production across Pakistan and northwestern India some 25 years after the first epidemic of the disease. Although resistance to the disease was introduced in the late 1990s, this succumbed quite quickly with the appearance of a resistance breaking strain of the virus complex. The ongoing breeding efforts have identified some sources of tolerance to the resistance breaking strain, but the desired immunity remains elusive [1]. Although the genetic changes in the complex pre- and post-resistance breaking have been determined, the precise mechanism of (molecular basis for) resistance breaking remains unresolved [35]. The betasatellite associated with CLCuD in south Asia, CLCuMB, has been shown to encode the major symptom determinant of the complex [19] and showed significant differences post-resistance breaking, specifically recombination within the SCR [24], which has been used as a genetic marker of the resistance-breaking virus complex.

Cotton is grown widely across the warmer parts of the world. CLCuD occurs in Asia and Africa with distinct etiologies, although in both areas begomovirus-betasatellite complexes are involved [1]. In all areas where cotton is grown the requirements for establishment of CLCuD are present—the main requirement being the presence of the vector *B. tabaci*. There is thus the fear that CLCuD could spread. In at least one instance this fear has been realized, with CLCuD reported for the first time in southeastern China [36]. The study here has assessed differences between the CLCuMB prevalent in two regions of Pakistan (Punjab and Sindh), following a recent upsurge of CLCuD problems in Sindh, and relates this to sequences changes in CLCuMB in the wider region.

For the CLCuD affected plants assessed here, CLCuMB was identified in all instances. This is consistent with the earlier demonstration of the importance of this specific betasatellite for causing the disease in cotton in south Asia [5,37]. The virus, however, appears less important. Experimentally even viruses not usually identified in cotton or associated with CLCuMB can induce CLCuD symptoms in cotton in the presence of CLCuMB—albeit transiently [38]. However, in the absence of CLCuMB even the viruses that do commonly infect cotton show poor infectivity to and atypical symptoms in cotton [5].

For CLCuMB, the sequences between the SCR and the A-rich region have previously been shown to be highly variable [24]. Deletion studies with *Ageratum* yellow vein betasatellite showed that the sequences between the SCR and the A-rich region are important for the *trans*-replication of the betasatellite by the helper begomovirus [39]. This finding was consistent with an earlier study of the *trans*-replication of an unusual betasatellite deletion mutant [40,41]. Nawaz-ul-Rehman *et al.* [41] showed that when *trans*-replicated *in planta* by *Cabbage leaf curl virus*, a virus that is not associated with a betasatellite, mutations occur in the sequences between the SCR and the A-rich region, and levels of betasatellite DNA are higher, suggesting that the sequence changes improve *trans*-replication. These studies thus indicate that betasatellite sequences between the SCR and the A-rich region are likely constrained by, and change due to, interaction with the helper virus. The changes in these sequences of CLCuMB detected here, with the range of variants in Sindh showing commonalities and differences with both the Punjab and India, are consistent with a greater diversity of helper begomoviruses detected in cotton in Sindh than in the Punjab [23,25,27].

Analysis of the predicted amino acid sequences of the  $\beta$ C1 gene of CLCuMB show that sequence variation in this protein is highly constrained.  $\beta$ C1 is the only gene encoded by betasatellites and

mediates all betasatellite functions including suppression of gene silencing, virus movement in plants and modulating host gene expression. These functions are mediated by interaction with host factors, a number of which have been identified [42,43]. This thus indicates that variation in  $\beta$ C1 is constrained by the need to maintain interactions with the host rather than the helper virus.

Analysis of the amino acid sequences of  $\beta$ C1 also seem to indicate that there is no correlation between resistance breaking and  $\beta$ C1 sequence. For example, CLCuMB-[PK:Bur1:02] (AM084379), a “Burewala strain” isolate, lacks the amino acid sequence changes present in the majority of CLCuMB<sup>Bur</sup>, suggesting that these sequence changes are not required for resistance breaking. This is consistent with the hypothesis, put forward by Amrao *et al.* [23], that the avirulence determinant for the CLCuD resistance in cotton introduced during the late 1990s was the C2 protein—based upon the finding that the virus involved in resistance breaking (CLCuBuV) lacks an intact C2 gene. Nevertheless, there are distinct differences between the  $\beta$ C1 sequences of CLCuMB occurring in the Punjab and Sindh for the majority of isolates. The reason for this is unclear, but could be due to the founder effect, with sequence changes selected for due to a requirement other than resistance breaking, spreading throughout a region. This in turn would suggest that the selection pressures in the Punjab and Sindh differed, the difference possibly being the distinct cotton varieties in the two regions. In light of the possibility of the betasatellite not being directly involved in resistance breaking in cotton, it would seem expedient to, in future, use CLCuBuV (specifically the truncated C2 gene; [23]) as the genetic marker of the resistance breaking virus complex rather than CLCuMB<sup>Bur</sup> (specifically the ToLCB recombination).

The function of the SCR of betasatellites remains a mystery. The position of the SCR in betasatellites is analogous to the position of the common region (CR) of bipartite begomoviruses [7]. The CR is a sequence shared between the two components of the genomes of bipartite begomoviruses which sits in the intergenic region, encompasses the promoter driving complementary-sense gene expression, sequence motifs (known as iterons; [44]) that are high affinity binding sites for the virus encoded replication-associated protein (Rep; a rolling-circle replication-initiator protein; [15]) as well as a predicted hairpin structure (containing the nonanucleotide sequence TAATATTAC, as part of the loop, that is nicked by Rep to initiate rolling-circle replication; [15]). The SCR also sits in an intergenic region but does not contain promoter elements, these instead being situated in the sequence upstream of the A-rich sequence ahead of the complementary sense  $\beta$ C1 gene [45], or iterons. The iterons or rather, iteron-like sequences, instead appear to be situated in the sequence between the SCR and the A-rich region [39,41]. The SCR does contain a predicted hairpin structure with a nonanucleotide sequence which presumably is required for the helper virus-encoded Rep to initiate satellite DNA replication, although this has not yet been proven experimentally. The position of the SCR, surrounding the origin of replication, may suggest a role in interaction with host factors involved in DNA replication. In light of this, it is difficult to know what benefits, if any, there are from recombination between betasatellites within the SCR, such as those for CLCuMB<sup>Bur</sup> and CLCuMB<sup>Sha</sup> other than to suggest that this may improve betasatellite trans-replication.

Although it is possible that each of these distinct CLCuMB types/strains has arisen independently by recombination (inserting ToLCB sequence and possibly also sequences of unknown origin into the SCR and also between SCR and A-rich region), a far more plausible explanation is that after an initial (large) insertion, the sequence of the recombinant fragment was sequentially reduced by recombination with CLCuMB<sup>Mul</sup>. Since resistance breaking, CLCuMB<sup>Mul</sup> has not been encountered in cultivated

cotton in the Punjab, appears not to be present in Sindh and has recently become rare in cultivated cotton in India (an example being CLCuMB-[IN:Bih:10] HM461864). Nevertheless, this strain of CLCuMB remains in the environment in weeds, non-cultivated cotton species and probably other crops such as tomato [46,47]. These sequence changes are thus more likely to be occurring in a host species other than cotton.

The finding that many of the features of the CLCuMB isolates from Sindh are shared with CLCuMB isolates from India (the full-length fragment upstream of the ToLCB sequences and the significant changes in the sequences between the SCR and the A-rich region), but not with isolates from the Punjab, suggests that there is more movement of CLCuMB, and thus presumably also the virus(es), between north western India and Sindh than previously thought. Why CLCuMB in the Punjab should be distinct is unclear. Possibly, the widespread adoption of resistant cotton varieties in the Punjab, which was not the case in either Sindh or India, has selected for distinct variants/strains of the satellite. In light of the evidence suggesting that the satellite is not involved (directly) with resistance breaking, the selection mechanism on CLCuMB in the Punjab may be indirect, mediated by the virus. This is supported by the finding that the range of CLCuD-associated viruses in the three regions differ: no diversity in the Punjab (only CLCuBuV present in cotton [23]); some diversity in India (CLCuBuV and CLCuRaV identified [4]); and some diversity in Sindh (*Cotton leaf curl Gezira virus*, *Cotton leaf curl Kokhran virus* and *Cotton leaf curl Shahdadpur virus* [25,27]). If this is the case, then it is likely that the diversity of the satellite is limited by the virus, which in turn is limited by the cotton variety.

The study presented shows that the CLCuD betasatellite in south Asia is evolving quite rapidly, although the forces driving this are unclear. A better understanding of virus-satellite interactions is required before a full understanding of the effects will be forthcoming. In addition the results have highlighted a high gene flow between the three regions—Punjab, Sindh and northwestern India. A better understanding of the movement of the pathogen might be useful in future efforts to control the disease; particularly should there be further epidemics.

## Acknowledgments

S.A., M.N.T. and S.J. were supported by the Higher Education Commission (HEC, Government of Pakistan) under the “Indigenous 5000 Fellowship Scheme”. G.R.B. was supported by Agriculture Department, Government of Sindh. R.W.B. was supported by the HEC under the “Foreign Faculty Hiring Scheme”.

## Author Contributions

S.A., M.N.T., G.R.B., S.J., A.Q.K. and I.A. performed experiments and prepared the manuscript. S.M. provided overall directions regarding the designing of all experiments, writing and supervised the work. R.W.B. was involved in critical review of the work and writing the manuscript. The final manuscript was read and approved by all authors.

## Conflicts of Interest

The authors declare no conflict of interest.

## References and Notes

1. Sattar, M.N.; Kvarnheden, A.; Saeed, M.; Briddon, R.W. Cotton leaf curl disease—An emerging threat to cotton production worldwide. *J. Gen. Virol.* **2013**, *94*, 695–710.
2. Briddon, R.W.; Markham, P.G. Cotton leaf curl virus disease. *Virus Res.* **2000**, *71*, 151–159.
3. Mansoor, S.; Briddon, R.W.; Zafar, Y.; Stanley, J. Geminivirus disease complexes: An emerging threat. *Trends Plant Sci.* **2003**, *8*, 128–134.
4. Rajagopalan, P.A.; Naik, A.; Katturi, P.; Kurulekar, M.; Kankanallu, R.S.; Anandalakshmi, R. Dominance of resistance-breaking *Cotton leaf curl Burewala virus* (CLCuBuV) in northwestern India. *Arch. Virol.* **2012**, *157*, 855–868.
5. Briddon, R.W.; Mansoor, S.; Bedford, I.D.; Pinner, M.S.; Saunders, K.; Stanley, J.; Zafar, Y.; Malik, K.A.; Markham, P.G. Identification of DNA components required for induction of cotton leaf curl disease. *Virology* **2001**, *285*, 234–243.
6. Mansoor, S.; Briddon, R.W.; Bull, S.E.; Bedford, I.D.; Bashir, A.; Hussain, M.; Saeed, M.; Zafar, M.Y.; Malik, K.A.; Fauquet, C.; *et al.* Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA  $\beta$ . *Arch. Virol.* **2003**, *148*, 1969–1986.
7. Brown, J.K.; Fauquet, C.M.; Briddon, R.W.; Zerbini, M.; Moriones, E.; Navas-Castillo, J. Geminiviridae. In *Virus Taxonomy—Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Inc.: London, UK; Waltham, MA/San Diego, CA, USA, 2012; pp. 351–373.
8. Melgarejo, T.A.; Kon, T.; Rojas, M.R.; Paz-Carrasco, L.; Zerbini, F.M.; Gilbertson, R.L. Characterization of a New World monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. *J. Virol.* **2013**, *87*, 5397–5413.
9. Sánchez-Campos, S.; Martínez-Ayala, A.; Márquez-Martín, B.; Aragón-Caballero, L.; Navas-Castillo, J.; Moriones, E. Fulfilling Koch's postulates confirms the monopartite nature of tomato leaf deformation virus, a begomovirus native to the New World. *Virus Res.* **2013**, *173*, 286–293.
10. Briddon, R.W.; Mansoor, S. Beta ssDNA satellites. In *Encyclopedia of Virology*; Mahy, B.W.J., van Regenmortel, M.H.V., Eds.; Academic Press: Oxford, UK, 2008; pp. 314–321.
11. Briddon, R.W.; Ghabrial, S.; Lin, N.S.; Palukaitis, P.; Scholthof, K.B.G.; Vetten, H.J. Satellites and other virus-dependent nucleic acids. In *Virus Taxonomy—Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Inc.: London, UK; Waltham, MA/San Diego, CA, USA, 2012; pp. 1209–1219.
12. Briddon, R.W.; Bull, S.E.; Amin, I.; Mansoor, S.; Bedford, I.D.; Dhawan, P.; Rishi, N.; Siwath, S.S.; Abdel-Salam, A.M.; Markham, P.G. Diversity of DNA  $\beta$ ; a satellite molecule associated with some monopartite begomoviruses. *Virology* **2003**, *312*, 106–121.

13. Leke, W.N.; Sattar, M.N.; Ngane, E.B.; Ngeve, J.M.; Kvarnheden, A.; Brown, J.K. Molecular characterization of begomoviruses and DNA satellites associated with okra leaf curl disease in Cameroon. *Virus Res.* **2013**, *174*, 116–125.
14. Briddon, R.W. Cotton leaf curl disease, a multicomponent begomovirus complex. *Mol. Plant Pathol.* **2003**, *4*, 427–434.
15. Hanley-Bowdoin, L.; Settlege, S.B.; Orozco, B.M.; Nagar, S.; Robertson, D. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Crit. Rev. Plant Sci.* **1999**, *18*, 71–106.
16. Amin, I.; Hussain, K.; Akbergenov, R.; Yadav, J.S.; Qazi, J.; Mansoor, S.; Hohn, T.; Fauquet, C.M.; Briddon, R.W. Suppressors of RNA silencing encoded by the components of the cotton leaf curl begomovirus-betasatellite complex. *Mol. Plant Microbe Interact.* **2011**, *24*, 973–983.
17. Cui, X.; Li, G.; Wang, D.; Hu, D.; Zhou, X. A begomovirus DNA  $\beta$ -encoded protein binds DNA, functions as a suppressor of RNA silencing, and targets the cell nucleus. *J. Virol.* **2005**, *79*, 10764–10775.
18. Kon, T.; Sharma, P.; Ikegami, M. Suppressor of RNA silencing encoded by the monopartite tomato leaf curl Java begomovirus. *Arch. Virol.* **2007**, *152*, 1273–1282.
19. Qazi, J.; Amin, I.; Mansoor, S.; Iqbal, M.J.; Briddon, R.W. Contribution of the satellite encoded gene  $\beta$ C1 to cotton leaf curl disease symptoms. *Virus Res.* **2007**, *128*, 135–139.
20. Saeed, M.; Zafar, Y.; Randles, J.W.; Rezaian, M.A. A monopartite begomovirus-associated DNA  $\beta$  satellite substitutes for the DNA B of a bipartite begomovirus to permit systemic infection. *J. Gen. Virol.* **2007**, *88*, 2881–2889.
21. Iqbal, Z.; Sattar, M.N.; Kvarnheden, A.; Mansoor, S.; Briddon, R.W. Effects of the mutation of selected genes of *Cotton leaf curl Kokhran virus* on infectivity, symptoms and the maintenance of Cotton leaf curl Multan betasatellite. *Virus Res.* **2012**, *169*, 107–116.
22. Saunders, K.; Bedford, I.D.; Briddon, R.W.; Markham, P.G.; Wong, S.M.; Stanley, J. A unique virus complex causes *Ageratum* yellow vein disease. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6890–6895.
23. Amrao, L.; Amin, I.; Shahid, M.S.; Briddon, R.W.; Mansoor, S. Cotton leaf curl disease in resistant cotton is associated with a single begomovirus that lacks an intact transcriptional activator protein. *Virus Res.* **2010**, *152*, 153–163.
24. Amin, I.; Mansoor, S.; Amrao, L.; Hussain, M.; Irum, S.; Zafar, Y.; Bull, S.E.; Briddon, R.W. Mobilisation into cotton and spread of a recombinant cotton leaf curl disease satellite. *Arch. Virol.* **2006**, *151*, 2055–2065.
25. Amrao, L.; Akhter, S.; Tahir, M.N.; Amin, I.; Briddon, R.W.; Mansoor, S. Cotton leaf curl disease in Sindh province of Pakistan is associated with recombinant begomovirus components. *Virus Res.* **2010**, *153*, 161–165.
26. Zaffalon, V.; Mukherjee, S.; Reddy, V.; Thompson, J.; Tepfer, M. A survey of geminiviruses and associated satellite DNAs in the cotton-growing areas of northwestern India. *Arch. Virol.* **2011**, *157*, 483–495.
27. Tahir, M.N.; Amin, I.; Briddon, R.W.; Mansoor, S. The merging of two dynasties—Identification of an African cotton leaf curl disease-associated begomovirus with cotton in Pakistan. *PLoS One* **2011**, *6*, e20366.
28. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.

29. Briddon, R.W.; Bull, S.E.; Mansoor, S.; Amin, I.; Markham, P.G. Universal primers for the PCR-mediated amplification of DNA  $\beta$ ; a molecule associated with some monopartite begomoviruses. *Mol. Biotechnol.* **2002**, *20*, 315–318.
30. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
31. Page, R.D.M. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **1996**, *12*, 357–358.
32. Martin, D.P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefevre, P. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* **2010**, *26*, 2462–2463.
33. Briddon, R.W.; Brown, J.K.; Moriones, E.; Stanley, J.; Zerbini, M.; Zhou, X.; Fauquet, C.M. Recommendations for the classification and nomenclature of the DNA- $\beta$  satellites of begomoviruses. *Arch. Virol.* **2008**, *153*, 763–781.
34. Shahid, M.S.; Mansoor, S.; Briddon, R.W. Complete nucleotide sequences of cotton leaf curl Rajasthan virus and its associated DNA beta molecule infecting tomato. *Arch. Virol.* **2007**, *152*, 2131–2134.
35. Briddon, R.W.; Akbar, F.; Iqbal, Z.; Amrao, L.; Amin, I.; Saeed, M.; Mansoor, S. Effects of genetic changes to the begomovirus/betasatellite complex causing cotton leaf curl disease in South Asia post-resistance breaking. *Virus Res.* **2014**, in press.
36. Cai, J.H.; Xie, K.; Lin, L.; Qin, B.X.; Chen, B.S.; Meng, J.R.; Liu, Y.L. Cotton leaf curl Multan virus newly reported to be associated with cotton leaf curl disease in China. *Plant Pathol. J.* **2010**, *59*, 794–795.
37. Mansoor, S.; Amin, I.; Iram, S.; Hussain, M.; Zafar, Y.; Malik, K.A.; Briddon, R.W. The breakdown of resistance in cotton to cotton leaf curl disease in Pakistan. *Plant Pathol.* **2003**, *52*, 784.
38. Saeed, M. Tomato leaf curl virus and Cotton leaf curl Multan betasatellite can cause mild transient symptoms in cotton. *Australas. Plant Dis. Notes* **2010**, *5*, 58–60.
39. Saunders, K.; Briddon, R.W.; Stanley, J. Replication promiscuity of DNA- $\beta$  satellites associated with monopartite begomoviruses; deletion mutagenesis of the Ageratum yellow vein virus DNA- $\beta$  satellite localizes sequences involved in replication. *J. Gen. Virol.* **2008**, *89*, 3165–3172.
40. Lin, B.; Behjatnia, A.A.; Dry, I.B.; Randles, J.W.; Rezaian, M.A. High-affinity Rep-binding is not required for the replication of a geminivirus DNA and its satellite. *Virology* **2003**, *305*, 353–363.
41. Nawaz-ul-Rehman, M.S.; Mansoor, S.; Briddon, R.W.; Fauquet, C.M. Maintenance of an old world betasatellite by a new world helper begomovirus and possible rapid adaptation of the betasatellite. *J. Virol.* **2009**, *83*, 9347–9355.
42. Kumar, P.; Ushaa, R.; Zrachya, A.; Levyb, Y.; Spanov, H.; Gafni, Y. Protein–protein interactions and nuclear trafficking of coat protein and  $\beta$ C1 protein associated with bhendi yellow vein mosaic disease. *Virus Res.* **2006**, *122*, 127–136.
43. Yang, J.-Y.; Iwasaki, M.; Machida, C.; Machida, Y.; Zhou, X.; Chua, N.-H.  $\beta$ C1, the pathogenicity factor of TYLCCNV, interacts with AS1 to alter leaf development and suppress selective jasmonic acid responses. *Genes Dev.* **2008**, *22*, 2564–2577.

44. Argüello-Astorga, G.R.; Ruiz-Medrano, R. An iteron-related domain is associated to motif 1 in the replication proteins of geminiviruses: Identification of potential interacting amino acid-base pairs by a comparative approach. *Arch. Virol.* **2001**, *146*, 1465–1485.
45. Eini, O.; Dogra, S.; Selth, L.A.; Dry, I.B.; Randles, J.W.; Rezaian, M.A. Interaction with a host ubiquitin-conjugating enzyme is required for the pathogenicity of a geminiviral DNA  $\beta$  satellite. *Mol. Plant Microbe Interact.* **2009**, *22*, 737–746.
46. Azhar, M.T.; Akhtar, S.; Mansoor, S. Cotton leaf curl Multan betasatellite strains cloned from *Gossypium barbadense* further supports selection due to host resistance. *Virus Genes* **2012**, *45*, 402–405.
47. Nawaz-ul-Rehman, M.S.; Briddon, R.W.; Fauquet, C.M. A melting pot of Old World begomoviruses and their satellites infecting a collection of *Gossypium* species in Pakistan. *PLoS One* **2012**, *7*, e40050.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).