TITLE: **Genetic Diversity of *Cotton leaf crumple virus* in the Western Hemisphere**

DISCIPLINE: **Genomics and Biotechnology**

AUTHORS: Ali M. Idris
Department of Plant Sciences
University of Arizona
Tucson, AZ 85721

Judith K. Brown (Corresponding author)
Department of Plant Sciences
University of Arizona
Tucson, AZ 85721
Phone: 520-621-1402
Fax: 520-621-8839
jbrown@ag.arizona.edu

ACKNOWLEDGEMENT: This research project was supported in part by Cotton Incorporated.

Genetic Diversity of *Cotton leaf crumple virus* in the Western Hemisphere

ABSTRACT

Cotton leaf crumple virus **(CLCrV) a New World begomovirus of cotton and malvaceous weeds was detected in cotton, kenaf and cheeseweed exhibiting leaf crumple and stunting symptoms collected from filed in Guatemala, Mexico, and the US (AZ, CA and TX). A fragment containing the coat protein (CP) was amplified, cloned and sequenced. Phylogenetic analysis of the CP nucleotide sequence placed the CLCrV isolates in a clade containing *Sida golden mosaic Costa Rica virus*. The analysis partitioned the CLCrV complex into two groups, constituting geographically isolated genotypes. One genotype, the west coast, was identified in Arizona, California, Mexico and Guatemala, while the second genotype was identified in Texas. Although the genotype from Texas induces similar phenotype in cotton and is most closely related to the west coast genotype, it represents an example of begomovirus undergoing speciation.**

Begomoviruses cause widespread, damaging diseases of cotton in Old and New World locales. They are a large and diverse group of plant viruses (genus *Begomovirus*, family *Geminiviridae*) that infect a wide range of eudicots (http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/29000000.htm). They are transmitted in nature exclusively by the whitefly *Bemisia tabaci* (Genn.) complex (Brown, 2007).

The circular ssDNA genome of begomoviruses is arranged as a single component (monopartite), or as two separate ssDNA components (bipartite). The begomoviral genome encodes several multifunction proteins involved in rolling circle replication of the genome, gene transcription, cell-to-cell and long-distance movement, suppression of host gene silencing, and encapsidation of the viral genome (Lazarowitz, 1992).

Cotton (*Gossypium* sp.) is cultivated in diverse geographical locales in temperate and tropical zones. The crop is grown mainly for fiber production, however, cottonseed is used for oil extraction and cottonseed oil cake is processed into animal feed. In the Old World, the first cotton disease was reported in Nigeria in 1912, and subsequently, cotton leaf curl disease (CLCuD) was reported for the first time in Sudan during 1927. CLCuD is characterized by vein thickening, leaf curl and occasionally leaf enations (Idris and Brown, 2002). The causal agent of the CLCuD was shown to be whitefly-transmissible (Nour and Nour, 1964), which in retrospect suggested it was a whitefly-transmitted plant virus. The complete monopartite, ssDNA genome of the *Cotton leaf curl Gezira virus* (CLCuGV) was cloned and the sequence was determined (Idris and Brown, 2002). The disease is now known to be caused by a complex of infectious molecules that includes the 'helper' virus CLCuGV and a satellite DNA of the *beta* type (Idris et al., 2005) that has serves as a pathogenicity determinant by suppressing host-induced silencing of viral transcripts (Cui et al., 2005). Strains of CLCuGV and satellite DNAs also have been

discovered infecting several cultivated and wild species in the *Malvaceae* in Sudan and Egypt (Idris et al., 2002; Idris and Brown, 2002). Similar leaf curling symptoms were reported during a disease outbreak in Pakistan during 1980-81 in *G. hirsutum*. Based on the leaf curl symptom phenotype, which was similar to that reported in diseased cotton from Sudan, and evidence of whitefly transmission, the virus from Pakistan was also referred to as CLCuD and the causal agent as CLCuV. This leaf curl disease in Pakistan also is caused by a unique complex of DNA molecules that involves a monopartite begomovirus and a *beta* satellite DNA (Briddon et al., 2001). CLCuV and CLCuGV are now recognized as distinct, Old World begomovirus species.

In several New World locations, cotton production was reported to be negatively affected by a disease called cotton leaf crumple (CLCrD) during the 1950's. Characteristic symptoms were foliar crumpling and discoloration, and shortened internodes. The leaf crumple disease became widespread in irrigated cotton crops in Arizona and California, U.S.A. and in Sonora, the state in Mexico immediately to the south of Arizona (Dickson et al., 1954). Similar symptoms were reported in cotton in the south coast of Guatemala along with a disease exhibiting a distinct mosaic phenotype (Brown, 2002). In all cases the whitefly *B. tabaci* was associated with affected fields and was suspected as the vector of the causal agent.

The virus-like pathogen in Arizona U.S.A. was shown to be transmitted by the whitefly *B. tabaci* and to have geminate particle morphology when examined by transmission electron microscopy (Brown and Nelson, 1984). The genomic components of the *Cotton leaf crumple virus* (CLCrV were cloned and the DNA sequence for each was determined. The cloned genomic components of CLCrV were used to biolistically inoculate cotton seedlings and the symptoms characteristic of those observed in the field were reproduced, thereby corroborating begomovirus etiology of the disease (Idris and Brown, 2004).

In this study we have investigated the phylogenetic relationships based on the viral coat protein gene for CLCrV and apparent relatives from various geographic regions where cotton is grown in the New World. We have PCR-amplified, cloned, and determined the DNA sequenced for the coat protein (CP) encoded by the AV1 open reading frame (ORF), the most conservative ORF in begomoviruses. The CP sequence has been found to be a useful for provisional classification of begomoviruses, when the complete genome sequence is unavailable, and allows for certain key insights into virus origin and relatedness (Brown et al., 2001). Results provide a preliminary indication of genetic variability of CLCrV in Guatemala, Mexico and the Sunbelt of the US.

MATERIAL AND METHODS

Virus sources. Cotton leaf samples showing typical field symptoms of CLCrD (Fig. 1) were collected from Arizona, California, Texas, Guatemala, and Mexico. In addition, cheeseweed (*Malva parviflora*) showing leaf curling and stunting symptoms was collected from Maricopa County, AZ, while leaves from kenaf (*Hibiscus cannabinus*) showing leaf curling symptoms was collected from the Lower Rio Grande Valley, Texas. Symptoms in kenaf were widespread in the field by late-season at ~20% incidence. Total nucleic acids were extracted using the method of Doyle and Doyle (1987) and stored at –80C.

PCR amplification and sequencing. Total nucleic acid extracts from symptomatic leaves were subjected to PCR amplification using begomovirus-specific PCR degenerate primers, prAV2644 and prAC1154, flanking the coat protein gene (AV1) (Idris and Brown, 1998). For these isolates a 1100bp-fragment containing the CP was amplified, cloned into pGEM-T Easy (Promega), Clones carrying the PCR-amplified fragment were confirmed by screening minipreps with *Eco*R I digestion to release the expected size and then sequenced with

the universal primers T7 and SP6, having annealing sites on the plasmid vector. DNA sequences for each clone were used to assemble a contig using SeqMan software (DNASTAR Inc., Madison, WI). EditSeq software was used to search the contigs for the 756-nt ORF for AV1.

Phylogenetic analysis. The complete nucleotide sequence determined for each CLCrV CP was aligned with the CP sequence for a suite of begomoviruses selected from the GenBank database, using MegAlign software (DNASTAR Inc., Madison, WI). Those begomoviruses used in the analysis were based on BLAST search of GenBank database (data not shown) using CLCrV CP sequences obtained in this study. The sequences were corrected for accuracy by eye to maintain it in-frame. The aligned sequences were used to reconstruct phylogenetic trees using the maximum likelihood (ML) method available in PAUP* 4.0.0b10 (Swofford, 1998). ML trees were estimated using the GTR and gamma distribution options, and the SPR branch-swapping algorithm of the heuristic search method, with default settings for all other parameters.

RESULTS AND DISCUSSION

Fifteen CP sequences representing thirteen field samples collected from cotton or other malvaceous species in Guatemala, Mexico and the USA (AZ, CA, and TX) were obtained (Fig. 2). The CP sequences were all of the size (756 nt) expected for New World begomoviruses. Comparison between the field isolates from malvaceous hosts (cotton, kenaf and cheeseweed) with a suite of well-studied begomoviruses indicated the presence of a single begomovirus represented by field isolates collected from the New World despite distinct symptomatologies observed in these two regions (Fig. 2). The CLCrV isolates from Guatemala, Mexico, and the U.S. states of AZ and CA (hereafter referred to as ‘U.S. west coast’ isolates), and the virus isolates from Texas showed the greatest nt divergence, despite the observation that they cause similar symptoms in cotton. The CP sequences for these two latter groups shared ~94% nt

identity. The within group identity ranged from 97-99% for the west coast isolates, while isolates from Texas shared 99% identity. All begomovirus isolates examined from these distinct geographical locations clustered within the same New World begomovirus clade that contains CLCrV.

Therefore, confirming our earlier finding (Idris and Brown, 2004) CLCrV isolates are likely indigenous to each of their extant locales and were not a result of an introduction from the Old World. The isolates examined here comprise very tight sister clades, which cluster as minor groups within the major clade, and in accordance with their geographical origin i.e. from where there were collected. Somewhat surprisingly, the begomovirus isolates from cotton in AZ and CA were more closely related to cotton isolates from Mexico and Guatemala, compared to the Texas isolates from cotton and kenaf. This might be explain by the fact that Guatemala, the west coast of Mexico, AZ and CA are part of geographical region that is not interrupted by any natural barrier, whereas, this major western region in North America and extending to Central America is separated from the part of the U.S. where Texas is situated, by vast deserts and mountain ranges.

The phylogenetic analysis of the CP nt sequences placed the CLCrV isolates in a clade with the *Sida golden mosaic Costa Rica virus* (SiGMCRV), while the analysis of the whole DNA A component of three isolates (AZ, Son, TX) placed it in the SLCV clade (Idris and Brown, 2004). This tree topology change suggesting that the A component of CLCrV could have resulted from intermolecular recombination, suggesting that the CP of CLCrV, like that of other begomoviruses, is highly conserved in regions subjected to structural and/or functional constraints. Indeed, the CP of begomoviruses has been shown to be multifunctional gene. The product of this gene is required for virion assembly, host range (Ingham et al., 1995), whitefly-

mediated transmission (Briddon et al., 1990) and in some hosts, symptom formation (Gardiner et al., 1988).

Generally, geography has played a major role on the distribution of begomoviruses as there are divided into Old World and New World groups. Moreover, the Old World begomoviruses are further divided into smaller groups according to their geographical origin in the Far East, Southwest Asia, Africa, and North Africa and the Mediterranean region. In this study, we showed that even for a single virus species occurring in a zone of low diversity, such as CLCrV, its isolates group based on extant geographic origin and not with a strict basis by host. The kenaf isolate from Texas grouped phylogenetically with isolates from cotton, also collected in Texas, and similarly the cheeseweed isolate of CLCrV from Arizona clustered with western isolates from cotton (Fig. 2). It appears that the two genotypes of CLCrV that are extant in Central America and west Mexico have radiated into one of two regions, western U.S.A.-Mexico-Guatemala, and U.S.-TX. The natural host range of the CLCrV complex is narrow and has been found to naturally infect only field cotton, cheeseweed, kenaf, and common bean. The interaction between CLCrV populations and other begomoviruses apparently has become limited, possibly owing to their residing in locales in which begomovirus diversity is low and also owing to the exploitation of primarily cultivated species as hosts.

CLCrV is one of the few begomoviruses with a limited natural host range despite its rather wide distribution e.g. North and Central America. This criterion makes it an excellent candidate for studying population genetic variation. A large number of samples have been archived in our laboratory over the past twenty years. This makes possible not only the study of variation within and between CLCrV populations in cotton, but also of isolates associated with

other malvaceous species, for which it is now possible to evaluate the genetic variation of this cotton virus complex, toward greater evolutionary insights.

REFERENCES

Briddon, R.W., S. Mansoor, I.D. Bedford, M.S. Pinner, K. Saunders, J. Stanley, Y. Zafar, K. Malik, and P.G. Markham 2001. Identification of DNA components required for induction of cotton leaf curl disease. Virology 285: 234-243.

Briddon R.W., M.S. Pinner, J. Stanley, and P.G. Markham. 1990. Geminivirus coat protein gene replacement alters insect specificity. Virology 177: 85-94.

Brown, J.K. 2002. Plant Virus and Mycoplasma Diseases of Cotton. Cotton Compendium. Amer. Phytopath. Soc., St. Paul, MN.

Brown, J.K. 2007. The *Bemisia tabaci* complex: genetic and phenotypic variability drives begomovirus spread and virus diversification. Plant Disease APSNet Feature Article December-January 2006-07 [Online]. Available at http://www.apsnet.org/online/feature/btabaci/.

Brown, J.K., A.M. Idris, E. Torres-Jerez, G.K. Bank, and S.D. Wyatt. 2001. The core region of the coat protein gene is highly useful for establishing the provisional identification and classification of Begomoviruses. Archive of Virology 146: 1581-1598.

Brown, J.K., and M.R. Nelson. 1984. Geminate particles associated with cotton leaf crumple disease in Arizona. Phytopathology 71: 699-703.

Cui, X., G .Li, D. Wang, D. Hu, and X. Zhou. 2005. A begomovirus DNAB-encoded protein binds DNA, functions as a suppressor of RNA silencing and target the cell nucleus. J. Virol. 79: 10764-10775.

Dickson, R.C., M.M. Johnson, and E.F. Laird. 1954. Leaf crumple, a virus disease of cotton. Phytopathology 44: 479-480.

Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bull. 19: 11-15.

Gardiner, W.E., G. Sunter, L. Brand, J.S. Elmer, S.G. Rogers, and D.M. Bisaro. 1988. Genetic analysis of tomato golden mosaic virus: The coat protein is not required for systemic spread or symptom development. EMBO Journal 7: 899-904.

Idris, A.M., R.W. Briddon, S.E. Bull, and J.K. Brown. 2005. Cotton leaf curl Gezira virus-satellite DNAs represent a divergent, geographically isolated Nile Basin lineage: predictive identification of a satDNA REP-binding motif. Virus Res. 109: 19-32.

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

Idris, A.M., and J.K. Brown. 2002. Molecular analysis of *Cotton leaf curl virus*-Sudan reveals an evolutionary history of recombination. Virus Genes 24: 249-256.

Idris, A.M., and J.K. Brown. 2004. *Cotton leaf crumple virus* is a distinct New World begomovirus species with complex evolutionary relationships indicative of recombination and reasortants. Phytopathology 94: 1068-1074.

Idris, A.M., M.H. Hussein, A.M. Abdel-Salam, and J.K. Brown. 2002. Phylogenetic relationships for okra leaf curl-and hollyhock leaf crumple-associated begomoviruses and first report of associated satDNAs. Arab J. Biotech. 5: 67-82.

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

Lazarowitz, S.G. 1992. Geminiviruses: genome structure and gene function. Critical Rev. in Plant Sci. 11: 327-349.

Nour, A.M., and J.J. Nour. 1964. Identification, transmission and host range of cotton leaf curl viruses infecting cotton in the Sudan. Empire Cotton Growing Review 41: 27-37.

Swofford, D.L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Association.

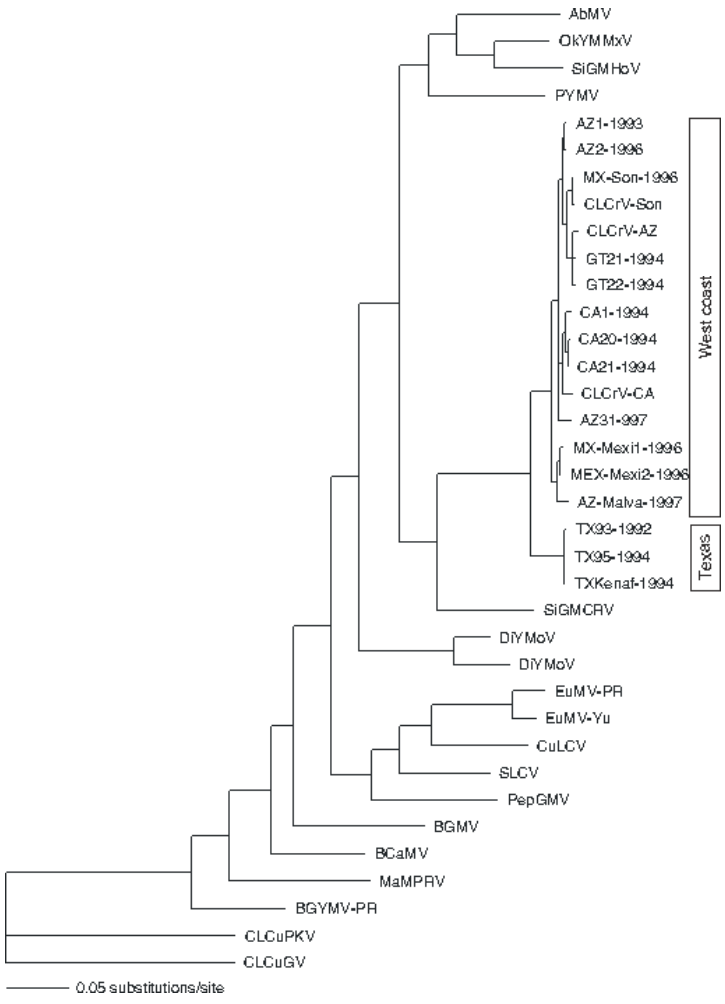
FIGURE CAPTIONS

Figure 1. Cotton seedling (*Gossypium hirsutum* L.) 'Delta Pine 5415' showing typical symptoms of *Cotton leaf crumple virus*-Arizona isolate.

Figure 2. Phylogenetic tree showing the predicted relationships the for nucleotide sequence for the coat protein gene for *Cotton leaf crumple virus* (CLCrV) and related begomoviruses. The tree was constructed using the heuristic search and maximum likelihood method in PAUP 4.0b version 10. The tree was rooted using the Old World *Cotton leaf curl Gezira virus* (CLCuGV).

Figure 1.

Figure 2.


AbMV
OkYMMxV
SiGMHoV
PYMV
AZ1-1993
AZ2-1996
MX-Son-1996
CLCrV-Son
CLCrV-AZ
GT21-1994
GT22-1994
CA1-1994
CA20-1994
CA21-1994
CLCrV-CA
AZ31-997
MX-Mexi1-1996
MEX-Mexi2-1996
AZ-Malva-1997
West coast
TX93-1992
TX95-1994
TXKenaf-1994
Texas
SiGMCRV
DiYMoV
DiYMoV
EuMV-PR
EuMV-Yu
CuLCV
SLCV
PepGMV
BGMV
BCaMV
MaMPRV
BGYMV-PR
CLCuPKV
CLCuGV
0.05 substitutions/site