Molecular characterization and the potential use of begomovirus associated DNA 1 as a silencing-expression vector

A dissertation submitted to Quaid-i-Azam University, Islamabad in partial fulfillment of requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY

By

Muhammad Shafiq Shahid

National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad

and

School of Biotechnology Quaid-i-Azam University, Islamabad Pakistan (2010)
This Humble Effort is Dedicated
to
My Mother, Father
and
Other Family Members
ACKNOWLEDGEMENTS

All praises and thanks are to Almighty Allah, the compassionate, the merciful, whom I never heard “Nay” whenever I knocked upon His door. I offer my humblest thanks to the greatest social reformer, “The Holy Prophet Hazrat Muhammad” (PBUH) the entire source of guidance for humanity as a whole, forever.

I owe my profound gratitude to our worthy Director Dr. Zafar M. Khalid, NIBGE, Faisalabad, for providing me such a congenial environment to work in, during my PhD. Especially in such a undeveloped country, I was facilitated enough to carry out such a fine Research with support from Pakistan Atomic Energy Commission (PAEC) and Higher Education Commission (HEC) of Pakistan, the credit goes to former chairman Dr. Atta-ur-Rehman an inspiring scientist for our Nation to always acknowledge.

My words will always remain incomplete for such a scientist who walks in the lab with new trends in science, Dr. Shahid Mansoor, Head Agricultural Biotechnology Division NIBGE. He taught me what in actual sense innovation in science does. I always wish to carry on the same attitude and spirit to establish myself for the betterment of science.

The best part of my PhD Research lies with an opportunity to work with one of the excellent Molecular Virologists, my worthy Supervisor Professor Dr. Robert William Briddon (Foreign Faculty Member and British Scientist) Agricultural Biotechnology Division NIBGE. He trained me what I didn’t know earlier. I really can’t express my gratitude in actual sense. He gave me the confidence to describe things in the best and easiest way.
Special thanks to Professor Dr. Judith K. Brown (USA) for her all time available help, advices, encouragement and sincere cooperation during my six months stay at the University of Arizona USA. She was always a source of motivation for me during my stay at U of A.

I have respectful appreciation for Dr. Muhammad Ali Idris (USA) our lab Incharge at the U of A. He always had a very helping and encouraging attitude towards me. I want to thank him for his great contributions, continuous encouragement, technical and moral support during my stay at U of A.

I personally acknowledge the help, guidance, and contributions of Mr. Imran Amin and Mrs. Javaria Qazi. A special thank to my friends and Colleagues, they are always a source of inspiration for me.

Last but not the least; I am indebted to my mother, for her infinite prayers and love, My brothers Ch. M. Rafiq, Walayat Ali, Liaquat Ali, Majid Ali, Sisters Sumaira Inayat, Sadia Inayat, Bhabies Asia Rafiq, Farah Walayat and all my Nephews and Nieces for their affection, care and devotion during my whole study.

Muhammad Shafiq Shahid
Table of Contents

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Page#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Table of contents</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>List of figures</td>
<td>XII</td>
<td></td>
</tr>
<tr>
<td>List of tables</td>
<td>XIII</td>
<td></td>
</tr>
<tr>
<td>Acronyms and abbreviations</td>
<td>XIV</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>XX</td>
<td></td>
</tr>
</tbody>
</table>

Chapter 1 | Introduction and review of literature | 1 |
1.1 | Plant viruses | 1 |
1.2 | Geminiviruses | 1 |
1.3 | Genome organization of geminiviruses | 2 |
1.3.1 | Mastreviruses | 2 |
1.3.2 | Curtoviruses | 4 |
1.3.3 | Topocuvirus | 6 |
1.3.4 | Begomoviruses | 7 |
1.4 | Functions of geminivirus proteins | 9 |
1.4.1 | Replication associated protein (Rep) | 9 |
1.4.2 | Transcriptional activator protein (TrAP) | 11 |
1.4.3 | Replication enhancer protein (REn) | 12 |
1.4.4 | C4 protein | 12 |
1.4.5 | Coat protein (CP) | 13 |
1.4.6 | Pre-coat protein (AV2/V2) | 14 |
1.4.7 Nuclear shuttle protein (NSP) and movement protein (MP) 15

1.5 Satellites associated with geminiviruses 16

1.5.1 Betasatellite 17

1.5.2 Alphasatellite 19

1.6 Transmission of geminiviruses 21

1.7 Geminivirus-host interactions 22

1.8 Movement of geminiviruses 25

1.9 Impact of begomoviruses 26

1.9.1 Worldwide epidemics associated with begomoviruses 26

1.9.2 Impact of begomoviruses associated diseases in Pakistan 28

1.10 Recombination among begomoviruses 29

1.11 Favored features for recombination among geminiviruses 30

1.11.1 Bipartite genome 30

1.11.2 Multiple/mixed infections 31

1.11.3 Acquisition of satellite DNA components 32

1.11.4 Insect vector 32

1.12 Replication of geminiviruses 33

1.13 Natural plant defense against viruses 36

1.14 Control of plant diseases 37
1.14.1 Genetically engineered resistance against viruses 37
1.14.2 Pathogen derived resistance 38
1.14.3 Protein-mediated pathogen-derived resistance 38
1.14.4 RNA-mediated pathogen derived resistance 41
1.14.5 Non-pathogen derived resistance 42

1.15 Other approaches for disease resistance 42

1.16 RNA silencing 44

1.17 RNA silencing through the exogenous introduction of transgene 46

1.17.1 Post-transcriptional gene silencing (PTGS) 46
1.17.2 Small RNA molecules 47

Chapter 2 Material and Methods 51

2.1 DNA Manipulation 51

2.1.1 Sample collection 51
2.1.2 DNA extraction from plant samples 51
2.1.3 Quantification of DNA 52

2.2 Amplification of DNA 52

2.2.1 PCR amplification of DNA 52
2.2.2 Rolling-circle amplification (RCA) 52
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>Cloning of amplified DNA</td>
<td>53</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Cloning of PCR product</td>
<td>53</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Cloning of RCA product</td>
<td>53</td>
</tr>
<tr>
<td>2.4</td>
<td>Transformation of competent cells</td>
<td>54</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Transformation of heat shock <em>Escherichia coli</em> competent cells</td>
<td>54</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Transfection of competent <em>Agrobacterium tumefaciens</em> cells</td>
<td>54</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Plasmid isolation</td>
<td>54</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Digestion of plasmid DNA</td>
<td>55</td>
</tr>
<tr>
<td>2.5</td>
<td>DNA Analysis</td>
<td>56</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Agarose-gel electrophoresis</td>
<td>56</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Southern blot analysis</td>
<td>56</td>
</tr>
<tr>
<td>2.6</td>
<td>Purification of DNA</td>
<td>58</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Gel extraction and PCR product purification</td>
<td>58</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Phenol-chloroform treatment of DNA</td>
<td>59</td>
</tr>
<tr>
<td>2.7</td>
<td>Microbiological Techniques</td>
<td>59</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Preparation of heat shock competent <em>Escherichia coli</em> cells</td>
<td>59</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Preparation of electro competent <em>Agrobacterium tumefaciens</em> cells</td>
<td>60</td>
</tr>
</tbody>
</table>
2.7.3  *Agrobacterium*-mediated inoculation  60

2.7.4  Plant growth conditions  61

2.7.5  Sequencing and sequence analysis  61

2.7.6  Photography and image processing/manipulation  62

Chapter 3  Results  64

3.1  Introduction  64

3.2  Methodology  67

3. 2.1  Virus source and DNA extraction  67

3.2.2  Detection of alphasatellite  71

3.3  Results  71

3.3.1  Detection of alphasatellites in crops and weeds  71

3.3.2  Sequencing and sequence Analysis  72

3.3.2.1  Analysis of alphasatellites isolated from cotton  72

3.3.2.2  Analysis of alphasatellites isolated from weeds  73

3.3.3  Phylogenetic analysis of alphasatellite sequences  74

3.4  Discussion  80
CHAPTER 4  Results  84

4.1  Introduction  84

4.2  Methodology  89

4.2.1  Modification of the Cotton leaf curl Multan alphasatellite for use as a VIGS/expression vector  89

4.2.1.1  Plasmid construction  89

4.2.1.2  Origin and cloning of gene/gene fragments for silencing/expression  90

4.2.1.3  Production of constructs for *Agrobacterium*-mediated inoculation  90

4.3  Results  93

4.3.1  Production and testing of constructs  93

4.3.2  Alphasatellite vector-mediated silencing of GFP  94

4.3.3  Alphasatellite vector-mediated silencing of endogenous gene  95

4.4  Alphasatellite-mediated expression of GFP in plants  98

4.5  Discussion  99

Chapter 5  Results  103

5.1  Introduction  103

5.2  Methodology  104
5.2.1 Virus source

5.2.2 Cloning of begomovirus genomes and satellite DNAs

5.2.3 Production of constructs for Agrobacterium-mediated inoculation

5.2.4 Agro-infiltration and assessment of inoculated plants

5.3 Results

5.3.1 Sequence analysis

5.3.2 Alphasatellite

5.3.4 Infectivity to *N. benthamiana*

5.3.5 Infectivity to tomato

5.4 Discussion

Chapter 6 General Discussion

Chapter 7 References
<table>
<thead>
<tr>
<th>Figure#</th>
<th>Title</th>
<th>Page#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Typical genome organization of mastreviruses and the leafhopper</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><em>Cicadulina mbila</em> – vector of <em>Maize streak virus</em></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Typical genome organization of curtoviruses and <em>Circulifer tenellus</em></td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>Typical genome organization of <em>Tomato pseudo curly top virus</em> and</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>its vector <em>Micratus malleifera</em></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>Typical genome organization of begomoviruses and their vector</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td><em>Bemisia Tabaci</em></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>Genetic organization of betasatellites</td>
<td>20</td>
</tr>
<tr>
<td>1.6</td>
<td>Genetic organization of alphasatellites</td>
<td>20</td>
</tr>
<tr>
<td>1.7</td>
<td>Geminivirus DNA replication cycle showing intercellular movement of</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>viral DNA</td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>A model depicting geminivirus-induced RNA silencing in plants.</td>
<td>50</td>
</tr>
<tr>
<td>2.1</td>
<td>Capillary blotting apparatus for the transfer of nucleic acids from</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>agarose gels to Hybond membranes</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Symptoms exhibited by the crop and weed plants used for the alphasatellite study</td>
<td>67</td>
</tr>
<tr>
<td>3.2</td>
<td>Map of Pakistan showing the geographical origin of the plant samples used in the alphasatellite study.</td>
<td>68</td>
</tr>
<tr>
<td>3.3</td>
<td>Southern blots of alphasatellites from cotton</td>
<td>70</td>
</tr>
<tr>
<td>3.4</td>
<td>Phylogenetic tree of alphasatellites from cotton and weeds</td>
<td>77</td>
</tr>
<tr>
<td>4.1</td>
<td>Alphasatellite showed the strategy of the constructs</td>
<td>91</td>
</tr>
<tr>
<td>4.2</td>
<td>Silencing of GFP and ChlI in 16c and <em>N. benthamiana</em> by alphasatellite vectors</td>
<td>97</td>
</tr>
<tr>
<td>4.3</td>
<td>GFP expression of <em>N. benthamiana</em></td>
<td>99</td>
</tr>
<tr>
<td>5.1</td>
<td><em>N. benthamiana</em> plants showing symptoms with tomato leaf curl Oman complex</td>
<td>111</td>
</tr>
<tr>
<td>5.2</td>
<td>Tomato plants showing symptoms with tomato leaf curl Oman complex</td>
<td>113</td>
</tr>
<tr>
<td>5.3</td>
<td>Phylogenetic tree of <em>Tomato leaf curl Oman virus</em> (ToLCOMV) and</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td><em>Tomato yellow leaf virus Oman</em> (ToYLCV-OM)</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>Phylogenetic tree of alphasatellite from Oman complex</td>
<td>117</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table#</th>
<th>Title</th>
<th>Page#</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Details of the primers used during the study</td>
<td>63</td>
</tr>
<tr>
<td>3.1</td>
<td>Detection of alphasatellites from cotton and weeds</td>
<td>69</td>
</tr>
<tr>
<td>3.2</td>
<td>Features of alphasatellite molecules amplified from cotton and weeds</td>
<td>76</td>
</tr>
<tr>
<td>3.3</td>
<td>Highest and lowest percentage identity of alphasatellites from cotton and weeds with selected alphasatellites from the databases</td>
<td>79</td>
</tr>
<tr>
<td>4.1</td>
<td>Geminiviruses and associated components used as VIGS vectors</td>
<td>87</td>
</tr>
<tr>
<td>4.2</td>
<td>Plant viruses that have been used for VIGS study</td>
<td>88</td>
</tr>
<tr>
<td>4.3</td>
<td>Results of the inoculation of <em>N. benthamiana</em> plants with alphasatellite (and its derivatives) in the presence of helper viruses</td>
<td>92</td>
</tr>
<tr>
<td>5.1</td>
<td>Percent nucleotide identities for full-length genome, and open reading frames of ToLCOMV with selected monopartite begomoviruses</td>
<td>107</td>
</tr>
<tr>
<td>5.2</td>
<td>Inoculation results of <em>N. benthamiana</em> and tomato seedlings with components of the tomato leaf curl Oman complex</td>
<td>114</td>
</tr>
<tr>
<td>5.3</td>
<td>Highest and lowest percentage identity of Oman viruses with selected begomoviruses</td>
<td>117</td>
</tr>
<tr>
<td>5.4</td>
<td>Highest and lowest percentage identity of alphasatellites from the Oman complex with selected alphasatellites from the databases</td>
<td>119</td>
</tr>
</tbody>
</table>
# Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Full name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADK</td>
<td>Adenosine kinase</td>
</tr>
<tr>
<td>AYVV</td>
<td>Ageratum yellow vein virus</td>
</tr>
<tr>
<td>Avr</td>
<td>Avirulence</td>
</tr>
<tr>
<td>ACMV</td>
<td>African cassava mosaic virus</td>
</tr>
<tr>
<td>Bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BMCTV</td>
<td>Beet mild curly top virus</td>
</tr>
<tr>
<td>BCTV</td>
<td>Beet curly top virus</td>
</tr>
<tr>
<td>BGMV</td>
<td>Bean golden mosaic virus</td>
</tr>
<tr>
<td>BBTV</td>
<td>Banana bunchy top virus</td>
</tr>
<tr>
<td>BDMV</td>
<td>Bean dwarf mosaic virus</td>
</tr>
<tr>
<td>BYVMV</td>
<td>Bhendi yellow vein mosaic virus</td>
</tr>
<tr>
<td>CTAB</td>
<td>Cetyl triethyl ammonium bromide</td>
</tr>
<tr>
<td>CHS</td>
<td>Chalcone synthase</td>
</tr>
<tr>
<td>CLCuD</td>
<td>Cotton leaf curl disease</td>
</tr>
<tr>
<td>CLCuMV</td>
<td>Cotton leaf curl Multan virus</td>
</tr>
<tr>
<td>CLCuMB</td>
<td>Cotton leaf curl Multan betasatellite</td>
</tr>
<tr>
<td>CIAP</td>
<td>Calf intestine alkaline phosphate</td>
</tr>
<tr>
<td>CHL</td>
<td>Magnesium chelatase</td>
</tr>
<tr>
<td>CPMR</td>
<td>CP-mediated resistance</td>
</tr>
<tr>
<td>CP</td>
<td>Coat protein</td>
</tr>
<tr>
<td>CaLuCV</td>
<td>Cabbage leaf curl virus</td>
</tr>
<tr>
<td>CLCrV</td>
<td>Cotton leaf crumple virus</td>
</tr>
<tr>
<td>CaMV</td>
<td>Cauliflower mosaic virus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CLCuA</td>
<td>Cotton leaf curl alphasatellite</td>
</tr>
<tr>
<td>CMD</td>
<td>Cassava mosaic disease</td>
</tr>
<tr>
<td>CoYVV</td>
<td><em>Corchorus yellow vein virus</em></td>
</tr>
<tr>
<td>Ds</td>
<td>double stranded</td>
</tr>
<tr>
<td>DI</td>
<td>Defective interfering</td>
</tr>
<tr>
<td>DCL</td>
<td>Dicer-like proteins</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>EYVD</td>
<td><em>Eupatorium yellow vein disease</em></td>
</tr>
<tr>
<td>EMBL</td>
<td>European Molecular Biology Laboratory</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetracetic acid</td>
</tr>
<tr>
<td>EupYVV</td>
<td><em>Eupatorium yellow vein virus</em></td>
</tr>
<tr>
<td>EACMV</td>
<td><em>East African cassava mosaic virus</em></td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescence protein</td>
</tr>
<tr>
<td>FBNYV</td>
<td><em>Faba bean necrotic yellows virus</em></td>
</tr>
<tr>
<td>HR</td>
<td>Hypersensitive response</td>
</tr>
<tr>
<td>HCl</td>
<td>Hydrochloric acid</td>
</tr>
<tr>
<td>HLCuA</td>
<td>Hibiscus leaf curl alphasatellite</td>
</tr>
<tr>
<td>HYVMD</td>
<td><em>Honey suckle yellow vein mosaic disease</em></td>
</tr>
<tr>
<td>ICTV</td>
<td>International Committee on Taxonomy of Viruses</td>
</tr>
<tr>
<td>IR</td>
<td>Intergenic-region</td>
</tr>
<tr>
<td>ICMV</td>
<td><em>Indian cassava mosaic virus</em></td>
</tr>
<tr>
<td>ICAC</td>
<td>International Cotton Advisory Committee</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Dalton</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>LRR</td>
<td>Leucine-rich repeat</td>
</tr>
</tbody>
</table>
LIR Large intergenic region
MaYMoA Malvastrum yellow mosaic alphasatellite
MalYMHnA Malvastrum yellow mosaic Hainan alphasatellite
MYMV *Mungbean yellow mosaic virus*
MCS Multiple cloning site
MP Movement protein
mRNA messenger RNA
MDV *Milk vetch dwarf virus*
NCBI National Centre for Biotechnology Information
MSV *Maize streak virus*
NBS nucleotide-binding site
NLS Nuclear localization signal
NES Nuclear export signals
NIG NSP-interacting GTPase
Nt Nucleotide
NSP Nuclear shuttle protein
NaOH Sodium hydroxide
ORF Open reading frame
OYVMV *Okra yellow vein mosaic virus*
PTGS Post transcriptional gene silencing
PVX *Potato virus X*
PVY *Potato virus Y*
PDR Pathogen derived resistance
PHYVV *Pepper hausteco yellow vein virus*
PLRV *Potato leaf roll virus*
PAUP  Phylogenetic analysis using parimomy
PapLCV  *Papaya leaf curl virus*
PDS  Phytoene desaturase
PTR  Partial tandem repeat
pRBR  Plant retinoblastoma-like proteins
PCNA  Proliferating cell nuclear antigen
RCR  Rolling circle replication
Rep  Replication associated protein
REn  Replication enhancer protein
RNA  ribonucleic acid
RCA  Rolling circle amplification
ROS  Reactive oxygen species
RNAi  RNA interference
RdRp  RNA-dependent RNA-polymerase
SCSV  *Subterranean clover stunt virus*
SA  Salicylic acid
SAR  Systemic acquired resistance
SDS  Sodium dodecyl sulphate
SEL  Size exclusion limit
SACMV  *South African cassava mosaic virus*
SIR  Small intergenic region
SiGMV  *Sida golden mosaic virus*
SDW  Sterile distilled water
SCR  Satellite conserved region
SqLCV  *Squash leaf curl virus*
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCSV</td>
<td>Subterranean clover stunt virus</td>
</tr>
<tr>
<td>Ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>ToLCOMV</td>
<td>Tomato leaf curl Oman virus</td>
</tr>
<tr>
<td>TYLCVOM</td>
<td>Tomato yellow leaf curl virus Oman</td>
</tr>
<tr>
<td>TCV</td>
<td>Turnip crinkle virus</td>
</tr>
<tr>
<td>TIR</td>
<td>Toll-interleukin region</td>
</tr>
<tr>
<td>TLCD</td>
<td>Tomato leaf curl disease</td>
</tr>
<tr>
<td>TYLCD</td>
<td>Tomato yellow leaf curl disease</td>
</tr>
<tr>
<td>TMV</td>
<td>Tobacco mosaic virus</td>
</tr>
<tr>
<td>TRV</td>
<td>Tobacco rattle virus</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch virus</td>
</tr>
<tr>
<td>ToLCNDV</td>
<td>Tomato leaf curl New Delhi virus</td>
</tr>
<tr>
<td>ToLCMLV</td>
<td>Tomato leaf curl Mali virus</td>
</tr>
<tr>
<td>ToLCYTV</td>
<td>Tomato leaf curl Mayotte virus</td>
</tr>
<tr>
<td>TGMV</td>
<td>Tomato golden mosaic virus</td>
</tr>
<tr>
<td>TPCTV</td>
<td>Tomato pseudo curly top virus</td>
</tr>
<tr>
<td>ToLCV</td>
<td>Tomato leaf curl virus</td>
</tr>
<tr>
<td>TrAP</td>
<td>Transcriptional activator protein</td>
</tr>
<tr>
<td>TYLCV</td>
<td>Tomato yellow leaf curl virus</td>
</tr>
<tr>
<td>TYLCSV</td>
<td>Tomato yellow leaf curl Sardinia virus</td>
</tr>
<tr>
<td>ToMoV</td>
<td>Tomato mottle virus</td>
</tr>
<tr>
<td>TRSV</td>
<td>Tobacco ring spot virus</td>
</tr>
<tr>
<td>TGMV</td>
<td>Tomato golden mosaic virus</td>
</tr>
<tr>
<td>TYLCCNV</td>
<td>Tomato yellow leaf curl China virus</td>
</tr>
<tr>
<td>TYLCCNA</td>
<td>Tomato yellow leaf curl China alphasatellite</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>TbCSA</td>
<td>Tobacco curly shoot alphasatellite</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VIGS</td>
<td>Virus induced gene silencing</td>
</tr>
<tr>
<td>MVDV</td>
<td>Milk vetch dwarf virus</td>
</tr>
<tr>
<td>WDV</td>
<td>Wheat dwarf virus</td>
</tr>
</tbody>
</table>
Summary

The work presented in this thesis consists of three disparate elements which have the unifying feature of being concerned with alphasatellite. The satellite-like alphasatellites (previously known as DNA 1) were first identified in 1999 and were shown to be associated with the majority of begomoviruses (family Geminiviridae) that require the satellite molecule now collectively known as betasatellites. They are capable of autonomous replication in the cells of host plants, by virtue of encoding a rolling circle replication initiator protein (the replication associated protein [Rep]) but require a helper begomovirus for spread within and between plants. The begomovirus-betasatellite complexes from Asia have been shown to be invariably associated with alphasatellites. For all cotton leaf curl disease (CLCuD) complex affected plants from Pakistan and western India collected in the 1990s and early 2000s, which were examined for this component, an alphasatellite was detected. By PCR-mediated amplification and Southern blot hybridization it was shown that, for more recently collected samples, alphasatellites are no longer a constant companion of the CLCuD complex. However, although in the Punjab no alphasatellites were detected in CLCuD affected cotton, in Sindh some plants were found containing this molecule and sequence analysis showed these to represent new species. The situation in weeds associated with cotton fields and other crops was similar, with only a few plants containing alphasatellites. In this case the alphasatellites were shown to be both newly identified species and species that had previously been identified. The significance of these findings in relation to recent changes in the CLCuD complex and our present knowledge of the function(s) of alphasatellites is discussed.

Virus-induced gene silencing (VIGS) is a powerful reverse genetics tool for application in functional genomics. A VIGS vector, based on the Cotton leaf curl Multan alphasatellite (CLCuMA), was produced by deleting the A-rich region or by inserting a cloning site at the end of the Rep gene for subsequent insertion of DNA fragments of genes targeted for silencing. The alphasatellite-based VIGS vector was shown to be functional by silencing expression of GFP in GFP transgenic Nicotiana benthamiana (line 16c) plants and magnesium chelatase in N. benthamiana. By inoculation to plants with distinct helper begomoviruses it was shown that the vector
could be used for many, if not all begomoviruses, meaning that the vector could potentially be used in all plant hosts of begomoviruses; making this the VIGS vector with the broadest host range produced so far.

In addition, the alphasatellite vector was shown to be useful as a gene expression vector. A GFP containing vector was shown to be stably maintained in plants and to act as a marker of virus spread. This system will be useful for studying the movement of monopartite begomoviruses. For bipartite begomoviruses, which do not require the coat protein (CP) for infectivity and systemic spread, GFP expressing vectors have previously been produced using the CP-replacement approach. This system was used extensively to investigate the movement of the virus in planta. However, monopartite begomoviruses have an absolute requirement for the CP for infectivity, which preclude their study using the CP-replacement approach.

An analysis of tomato plants from Oman exhibiting severe leaf curl symptoms showed the presence of Tomato yellow leaf curl Oman (TYLCV-Om), a virus previously shown to be present in tomato in this country, a new recombinant begomovirus species, for which the name Tomato leaf curl Oman virus (ToLCOMV) is proposed, a betasatellite (tobacco leaf curl betasatellite (TbLCB) and an unusual alphasatellite. The alphasatellite (OM2) was shown to be an isolate of a previously identified satellite-like molecule associated with Ageratum yellow vein virus from Singapore and these are distinct from all other alphasatellites so far identified. Constructs for Agrobacterium-mediated inoculation were produced and inoculated to both Nicotiana benthamiana and tomato. This showed that the severe symptoms in plants are due to synergism between TYLCV-OM and ToLCOMV and that, in common with all previous studies of betasatellites, the presence of this molecule enhances symptoms. However, co-infection with OM2 attenuated symptoms. The attenuation was the most pronounced in the presence of the betasatellite and both viruses. The significance of these findings is discussed.
Chapter 1
1.1 Plant Viruses

There are currently three orders, 73 families, 9 subfamilies, 287 genera, and 1938 virus species (Fauquet et al., 2004). Most plant viruses have an RNA genome, although a minority of plant viruses have genomes of DNA. These viruses fall into two general types, those with circular double-stranded DNA (dsDNA), which replicate by reverse transcription through an RNA intermediate (the caulimoviruses and badnaviruses), and those with circular single-stranded DNA (ssDNA), which replicate through a dsDNA intermediate by a rolling circle mechanism (*Geminiviridae* and *Nanoviridae*) (Gutierrez, 1999; Jeske, Lutgemeier and Preiss, 2001).

1.2 Geminiviruses

The family *geminiviridae* includes a large number of plant-infecting viruses that produce, in many cases, very significant reductions in economically important crops of both monocotyledonous and dicotyledonous plants. This family of viruses is named after the unique geminate (twinned) virion morphology of its member viruses. Geminiviruses have either one or two circular ssDNA components of ~2.6–3.1 kb. (Stanley et al., 2005)

Geminiviruses fall into four genera, *mastrevirus*, *curtovirus*, *topocuvirus* and *begomovirus*, based on their genome structures, host ranges and insect vectors. The genus *mastrevirus* includes leafhopper-transmitted viruses with monopartite genomes that infect either monocots or dicots. *Maize streak virus* (MSV) and *Wheat dwarf virus* (WDV) are well-studied members of this genus. *Curtoviruses* are leafhopper transmitted viruses with monopartite genome that infect dicots. *Beet curly top virus* (BCTV) is one of the well-studied examples. Begomoviruses are dicot-infecting, whitefly transmitted viruses with either bipartite or monopartite genomes (Stanley et al., 2005). The most well known is *Bean golden yellow mosaic virus* (BGYMV; formerly called Bean golden mosaic virus, after which the genus is named). The genus *topocuvirus* has only a single member, *Tomato pseudo curly top virus* (TPCTV). This infects dicotyledonous plants and is transmitted by the treehopper *Micrulalis mallifera* (Simons and Coe, 1958).
The members of family *nanoviridae* have been divided into two genera based on their host ranges. The genus *babuvirus* consists of monocot-infecting viruses, the most well known of which is *Banana bunchy top virus*, an important pathogen of cultivated bananas. Viruses of the genus *nanovirus* infect dicotyledonous plants and include *Faba bean necrotic yellow vein virus*, *Milk vetch dwarf virus*, and *Subterranean clover stunt virus* (Gronenborn, 2004).

Geminiviruses are a major constraint to agricultural productivity in all tropical and sub-tropical regions of the world. Recently, they have spread into more temperate regions because of changes in agricultural practices and ecological conditions, as well as the global trade in agricultural products, which all encourage dissemination of the insect vectors. Most of the economically important diseases are caused by members of the genus *begomovirus* (Fauquet et al., 2003).

The genomes of geminiviruses range in from 2.5 to 5.6 kb (Stanley et al., 2005), which is among the smallest known genome for an independently replicating virus. They have an unusual, geminate capsid with two incomplete T=1 joined icosahedra. The capsids, range from 18-20 nm in diameter with a length of about 30 nm. The geminivirus group was established by the International Committee on the Taxonomy of Viruses (ICTV) in 1978 and upgraded to the family *geminiviridae* in 1995.

### 1.3 Genome organization of Geminiviruses

#### 1.3.1 Mastreviruses

Mastreviruses are monopartite geminiviruses, infecting either monocots or dicots. The viruses are obligately transmitted by leafhoppers (Homoptera: *Cicadellidae*), and are largely phloem-limited. These occur exclusively in the Old World. *Maize streak virus* (MSV) and *Wheat dwarf virus* (WDV) are the most well characterized members of the genus mastrevirus (Willment et al., 2007). Four proteins are encoded by mastrevirus genomes; RepA and Rep are encoded by C1 and a spliced product of the C1 and C2 ORFs, respectively, encoded on the complementary-strand while MP and CP by the V1 and V2 ORFs on virion strand (Palmer and Rybicki, 1998). Amongst geminiviruses, the Rep of mastreviruses is unique as its expression
involves splicing (Wright et al., 1997). In addition to its role in replication, Rep also regulates the expression of virion-sense genes whereas RepA is involved in cell-cycle control. MP and CP are involved in the movement and encapsidation of virus (Wright et al., 1997)

Specific interactions either between the mastrevirus MP and CP genes or their products are responsible for the infection efficiency, rate of symptom development and symptom severity (Van Der Walt et al., 2008; Wright et al., 1997) reported the presence of another intron in the V1 gene of MSV. This intron spans the region of V1 ORF that encodes the transmembrane domain of the movement protein of the mastreviruses. The presence of intron enhances the expression of coat protein gene expression (Boulton et al., 1993).

There are two intergenic regions located opposite to each other, a large and a small intergenic region (LIR and SIR, respectively) containing regulatory elements. Consensus promoter sequences for C1 and V1 ORFs have been found in the LIR (Palmer and Rybicki, 1998). The LIR also contains the origin of replication (ori) for virion-strand DNA synthesis similar to that of begomoviruses (Willment et al., 2007). The SIR contains bidirectional polyadenylation signals. The SIR also contains the ori for the complementary strand synthesis and a short ssDNA sequence (70-80 nts). This primer-like sequence is annealed to the encapsidated genomic ssDNA and is thought to prime the minus-strand synthesis.
**Introduction and Review of Literature**

![Image](image_url)

**Figure 1.1** *Cicadulina mbila*, the leafhopper vector of *Maize streak virus* and the typical genome organization of mastreviruses. The positions and orientations of genes are indicated. The genes are the movement protein (MP), the coat protein (CP) and the replication associated protein (Rep; see the text for the expression strategy of these two coding sequences).

### 1.3.2 Curtoviruses

In the genus *curtovirus*, *Beet curly top virus* (BCTV) is a well-studied example. Curtoviruses are dicot infecting monopartite viruses transmitted by leafhopper. The genomes of curtoviruses consist of one circular single-stranded DNA molecule of 2.9-3.0 kb (Baliji et al., 2004). Curtovirus are usually phloem-limited and viral replication, gene expression and virion formation occurs in the nucleus (Esau, 1977; Latham et al., 1997).

Recently an unusual geminivirus has been identified from Iran. This was named *Beet curly top Iran virus* and was tentatively placed in the genus *curtovirus* by the authors (Yazdi, Heydarnejad, and Massumi, 2008). Although it has the CP typical of viruses transmitted by *Circulifer tenellus*, the complementary-sense genetic arrangement is unusual. It encodes a typical Rep but has only one downstream ORF which may possibly be expressed by splicing with Rep (in many respects similar to the mastreviruses). This unusual arrangement suggests this virus may have a recombinant origin (or be very ancient – a possible progenitor of the curtoviruses).
which recombined with the begomoviruses to yield all extant curtoviruses). There have been suggestions in the geminivirus community that a new genus may need to be established to accommodate this virus. However, first Koch’s postulates will need to be satisfied and the gene expression strategy investigated.

Unlike other geminiviruses the intergenic region of curtoviruses is not sufficient to provide the full expression of C1 but transcriptional activator elements for C1 expression reside in the 3’ portion of C1 coding area itself and C1 protein does not auto-regulate its own expression (Hur et al., 2007). In order to identify the promoter motif involved in curtovirus sense-gene expression in transgenic Arabidopsis it was shown that curtovirus late gene expression by virion-sense promoters depends on the developmental stage of the host plant as well as on the number of conserved late element (CLE) motifs present in the promoter (Singh et al., 2008). In curtoviruses, a limited number of viruses infect a very wide range of plant species.

**Figure 1.2** *Circulifer tenellus*, the leafhopper vector of *Beet curly top virus* and the typical genome organization of curtoviruses. The positions and orientations of genes are indicated. The genes are the movement protein (MP), the coat protein (CP) and the replication associated protein (Rep) and the replication enhancer protein (REn). The precise functions of genes indicated as V2, C2 and C4 remain unclear.
1.3.3 Topocuvirus

Topocuviruses are treehopper-transmitted monopartite viruses with monopartite genomes that infect dicots. These are the only viruses as far as vector specificity is concerned, which are outside of *cicadellidae* (leafhoppers) and *aleyrodidae* (whiteflies). The only known member is *Tomato pseudo-curly top virus* (ToPCTV) (Briddon et al., 1996). The ToPCTV genome contains features typical of both mastreviruses and begomoviruses, showing that it is a natural recombinant. In line with this hypothesis, ToPCTV can trans-complement the movement of the DNA-A components of two bipartite begomoviruses, in the absence of their corresponding DNA-B (Briddon and Markham, 2001).

![Micrutalis malleifera](image1.png)

**Figure 1.3** *Micrutalis malleifera*, the treehopper vector of *Tomato pseudo curly top virus* and the genome organization of this virus. The positions and orientations of genes are indicated. The genes are the coat protein (CP) and the replication associated protein (Rep), deduced from similarities with the viruses of the other genera. The functions of genes indicated as V2, C2, C3 and C4 remain unclear.
1.3.4 Begomoviruses

The genus *begomovirus* contains more than 230 species (Fauquet et al., 2008). As a group they have a very wide host range, but infect only dicotyledonous plants. Worldwide they are responsible for a large amount of economic damage to many important crops such as tomatoes, beans, squash, cassava and cotton. The viruses are obligately transmitted by an insect vector, the whitefly *Bemisia tabaci* (*Gennadius*) (*Homoptera: Aleyrodidae*). Its efficiency as a vector is attributed to its polyphagous nature, although some biotypes can be very host specific. Begomoviruses have either monopartite or bipartite genomes. Affected plants exhibit a range of symptoms such as leaf curling, stunted growth and a poor yield (Briddon, 2003).

Many begomoviruses have bipartite genomes, the components of which are designated DNA A and DNA B (Fig. 1.1; Stanley and Gay, 1983; Howarth et al., 1985). These components share no sequence identity except for a highly conserved sequence known as common region (CR; Harrison, 1985). The common region contains the origin of virion-strand DNA replication and thus maintains the integrity of bipartite genomes (Hanley-Bowdoin, 1999).

Proteins required for viral DNA replication, encapsidation and the control of viral gene expression are encoded by DNA A (Townsend, Watts, and Stanley, 1986; Etessami, Watts, and Stanley, 1989) while DNA B encodes proteins required for intercellular and intracellular viral movement and symptom development (Von Arnim and Stanley, 1992; Sanderfoot and Lazarowitz, 1996; Bisaro, 1996). The components share a CR that contains motifs required for the control of gene expression and replication, notably conserved iteron motifs and a putative stem loop structure containing the highly conserved nonanucleotide (TAATATTAC) that are the origin of virion strand replication (Hanley-Bowdoin, 1999).

The DNA A component contains six open reading frames (ORFs), four (the replication associated protein (Rep), the transcriptional activator protein (TrAP), replication enhancer protein (REn), and AC4) on the complementary-sense strand and two, (coat protein (CP) and precoat-protein (AV2), on the virion-sense strand. Rep and REn are involved in replication since mutations of Rep blocked viral replication,
whereas REn helps in replication process as mutation in REn greatly reduced DNA levels and resulted in severely delayed and attenuated symptoms (Elmer et al., 1988; Etessami et al., 1988; Etessami, Watts, and Stanley, 1989; Morris et al., 1991). The CP is involved in the formation of a stable nucleoprotein complex for efficient spread in plants and insect transmission. The AV2 is absent in the New World begomoviruses.

The DNA B encodes two genes, the movement protein (MP) on the complementary-strand has a role in cell-to-cell movement (Noueiry, Lucas, and Gilbertson, 1994; Ward et al., 1997) and NSP on the virion-strand binds and transports ssDNA across the nuclear envelope (Pascal et al., 1994; Sanderfoot, Ingham, and Lazarowitz, 1996). Thus both components are essential for the establishment of efficient systemic infection.

The genomes of some begomoviruses from the Old World consist of a single DNA component (Dry et al., 1993; Kheyr-Pour et al., 1991; Mansoor, 1999; Navot et al., 1991; Noris, Accotto, and Luisoni, 1994). The genome of these viruses is homologous to the DNA A component of the bipartite viruses.

The majority of monopartite begomoviruses are associated with satellite and satellite-like molecules. These viruses are infectious to the host plants from which they were isolated but induce atypical symptoms /Ageratum yellow vein virus (AYVV) from Singapore and Cotton leaf curl Multan virus (CLCuMV) from Pakistan are such examples (Saunders et al., 2002; Briddon and Markham, 2000).

The first of these is a satellite-like molecule, collectively known as alphasatellites. These are approximately 1380 nucleotides in length and comprise a group of closely related molecules that encode a rolling-circle replication initiator protein (Rep) which shows high levels of sequence identity to the Reps encoded by components of nanoviruses (Mansoor, 1999; Briddon and Stanley, 2006). Alphasatellites depend upon helper begomoviruses for movement in plants, transmission between plants and apparently play no part in the etiology of the disease. Alphasatellite can replicate autonomously by virtue of encoding a Rep protein (a rolling-circle replication initiator protein).
1.4 Functions of Geminivirus proteins

1.4.1 Replication associated protein (Rep)

Rep, also known as C1, AC1 and AL1, is a multifunctional protein and the only viral protein absolutely required for virus replication. AC1 is encoded on the complementary sense strand. This protein is involved in several biological processes: initiation and termination of rolling circle replication (RCR) by nicking and religating the replication origin of viral DNA (Laufs et al., 1995) and repression of its own gene transcription (Eagle, Orozco, and Hanley-Bowdoin, 1994). The AC1 proteins of geminiviruses are closely related and show substantial sequence conservation. Four functional domains have been delineated for begomovirus Rep: the N-terminal domain (amino acids 1 to 120), which is involved in initiation by geminiviruses (Orozco et al., 2000), AC1 protein initiates rolling circle replication by a site-specific
Introduction and Review of Literature

cleavage within the loop of the conserved nonamer sequence, TAATATTAC (Heyraud-Nitschke et al., 1995).

The AC1 protein binding site is located between the TATA box and the transcription start site for the Rep gene and acts as the origin recognition sequence and as a negatively regulatory element for AC1 gene transcription (Eagle, Orozco, and Hanley-Bowdoin, 1994), the oligomerization domain (121 to 180 aa), leading to interactions with itself (Orozco et al., 2000) and with host factors (Hanley-Bowdoin, Settlage, and Robertson, 2004).

The mutations in its oligomerization domain affect both replication and AC1-mediated transcription repression (Orozco et al., 2000): the ATPase domain (amino acids 181 to 330), which is characterized by the presence of a P loop and a carboxyl-terminal domain (amino acids 331 to 359) of unknown function but shown to be required for viral replication in the case of Tomato golden mosaic virus (TGMV) (Orozco et al., 2000). The ATPase domain of geminivirus Rep proteins was identified as a common element among proteins encoded by small DNA and RNA viruses (Gorbalenya, Koonin, and Wolf, 1990) and is characterized by three conserved motifs: Walker A in the P loop, Walker B, and motif C.

The AC1 protein alone can initiate RCR without requiring other accessory viral factors (Hong, Stanley, and van Wezel, 2003). AC1 protein also has DNA helicase activity which depends upon the oligomeric state of the protein (Clerot and Bernardi, 2006). During rolling-circle replication AC1 binds to specific repeated sequences (known as iterons) present in the intergenic region (IR) and hydrolyses the phosphodiester bond between the seventh and eighth residues of the invariant nonamer 5' TAATATTAC 3' (Stanley, 1995; Laufs et al., 1995).

Recently it has been found that AC1 binds to the iteron sequences in a highly ordered manner and many molecules together are attached with the iteron sequences, which is required to destabilize the origin of replication (Singh et al., 2008). Protoplast studies have determined that repression of upstream transcription by AC1 protein enhances AC2 and AC3 expression in TGMV (Shung and Sunter, 2007). For
efficient replication, AC1 is assisted by a replication enhancer (REn; also called C3 or AL3) (Settlage, See, and Hanley-Bowdoin, 2005).

### 1.4.2 Transcriptional activator protein (TrAP)

TrAP is also known as AC2, C2 an AL2. AC2 is a ~15-KD a transcriptional activator protein unique to begomoviruses because it is absent in mastreviruses and a related protein in curtoviruses, AC2 protein, seems to play a different role. In mastreviruses AC1 protein provide the functions of AC2 (Liu, Davies, and Stanley, 1998).

TrAP is necessary for transactivation of late genes (Sunter and Bisaro, 1992; Sunter and Bisaro, 1997; Gröning, Hayes, and Buck, 1994; Jeffrey, Pooma, and Petty, 1996). Recently, people have shown that the AC2 gene of *Cabbage leaf curl virus* (CaLCuV) activates the CP promoter in mesophyll and acts to derepress the promoter in vascular tissue, similar to that observed for TGMV (Lacatus, 2008).

AC2 interacts with itself and localizes to the nucleus. Thus, AC2 self-interaction correlates with nuclear localization and efficient activation of transcription (Yang et al., 2007). Consistent with its function as a transcriptional activator, three conserved domains have been recognized in this protein: a basic domain with a nuclear localization signal (NLS) at the N terminus, a central DNA-binding domain with a non classical Zn-finger motif, and an acidic activator domain at C-terminus (Hartitz, Sunter, and Bisaro, 1999). A comparison of steady state transcript level and the transcript level determined by nuclear run on assay showed that activation of AV1 and BV1 gene expression by the AC2 protein primarily occurs at the transcriptional level. For example, ACMV infection activated a transgene that was under the control of the coat protein promoter (Hong, Saunders, and Stanley, 1997). Mutational analysis of AC2 in begomoviruses showed that the mutation of this ORF prevented systemic movement of the virus, produced no capsid protein and accumulated reduced amount of ssDNA in the transient assay (Hayes and Buck, 1989; Sunter et al., 1990). TrAP function is not virus specific, as the AC2 gene products of TGMV, ACMV, TYLCV, *Texas pepper virus* (TPV) and *Squash leaf curl virus* (SqLCV) complement a TGMV AC2 mutant in tobacco protoplasts (Sunter, Stenger, and Bisaro, 1994). So, AC2 is functionally interchangeable among begomoviruses.
This absence of functional specificity suggests that either all begomovirus late promoters contain a common sequence element recognized by AC2 or AC2 interacts with cellular proteins common to all begomovirus plant hosts to effect transcriptional activation. Wang et al. (2005) showed AC2 to be an effective silencing suppressor and showed that TGMV and BCTV AC2 can suppress RNA silencing directed against a GFP reporter gene when silencing is induced by a construct expressing an inverted repeat of GFP RNA (dsGFP). Notably, TYLCCNV AC2 requires functional NLS and Zn-finger domains to suppress silencing (Dong et al., 2003). Further, since AC21-100 is as effective a suppressor as the full-length AC2 protein, activation and silencing suppression appear to be independent activities (Wang et al., 2005).

Hussain et al. (2007) demonstrated the role of TrAP in inhibiting a hypersensitive response (HR), the first such activity identified for a plant-infecting virus. Analysis of all ToLCNDV-encoded genes pinpointed the AC2 as the factor mediating the anti-HR effect. Deletion mutagenesis showed the central region of AC2, containing a zinc finger domain and nuclear localization signal, to be important in inhibiting the HR. More recently, Gopal et al. (2007) showed that AC2 of Bhendi yellow vein mosaic virus (BYVMV) is involved in transactivation and only mildly in suppression of gene silencing of monopartite begomoviruses viruses and not in transmission.

1.4.3 Replication enhancer protein (REn)

REn, also named as AC3/AL3. AC3 is a ~16 KD a protein in curtoviruses and in begomoviruses, which does not have counterpart in mastreviruses. AC3 mutations also produce a reduced level of accumulation of viral DNA in transient replication assays (Stanley et al., 1992).

The AC3 protein greatly enhances viral DNA accumulation of curtoviruses and begomoviruses (Elmer et al., 1988; Sunter et al., 1990) by interacting with Rep (Settlage, Miller, and Hanley-Bowdoin, 1996). Experimental observations suggested that AC3 protein might increase the affinity of Rep for the origin. Complementation studies revealed that AC3 could act on heterologous viruses (Sunter, Stenger, and Bisaro, 1994).
1.4.4 AC4/C4 protein

AC4 is highly variable among begomoviruses, which is expressed from an open reading frame (ORF) embedded in the Rep ORF. The expression of C4 protein in transgenic *N. benthamiana* produced virus-like symptoms and further confirmed its role in symptom development (Latham et al., 1997). Mutation of AC4 in the bipartite geminiviruses, ACMV and TGMV, resulted in wild type symptoms and no role could be ascribed to this ORF (Etessami, Watts, and Stanley, 1989; Elmer et al., 1988) AC4 of ACMV enhanced pathogenicity by increasing the levels of DNA accumulation of viruses and was also found to have anti-PTGS activity (Vanitharani, 2004).

1.4.5 Coat protein (CP)

Geminiviral capsids are composed of a single CP, encoded by the V1 gene or (also known as AV1), depending on the geminivirus (Zhang et al., 2001). For monopartite geminiviruses, CP is essential for systemic spread through the plant (Boulton et al., 1991; Briddon et al., 1989). For bipartite geminiviruses, it is not absolutely necessary for this task (Gardiner et al., 1988; Unseld, Frischmuth, and Jeske, 2004). It can still complement this NSP gene of DNA B (Ingham, Pascal, and Lazarowitz, 1995). It can bind ssDNA as well as double-stranded DNA (dsDNA) in a sequence-independent manner (Lazarowitz and Beachy, 1999). MSV CP also bound ss and ds viral DNA in a sequence non-specific manner. Besides the encapsidation function, CP is also required for transmission of the virus between the plants. The CP of the monopartite geminiviruses facilitates the transfer of infecting viral DNA into the host cell nucleus and is essential for systemic virus movement (Boulton, 1989; Lazarowitz et al., 1989; Liu et al., 1999; Woolston et al., 1989).

In contrast, bipartite begomoviruses may not require coat protein for systemic transmission (Gardiner et al., 1988; Pooma et al., 1996; Stanley and Townsend, 1986) although the disease symptoms are often attenuated and the onset of disease is delayed when plants are systemically infected with mutant (Hayes and Buck, 1989; Sanderfoot and Lazarowitz, 1996; Unseld, Frischmuth, and Jeske, 2004). The CP also determines the vector specificity (Briddon, 1990; Höfer et al., 1997; Höhnle et al., 2001) and protects the viral ssDNA from degradation during transmission by the
Insect vector (Azzam et al., 1994) or mechanical inoculation (Frischmuth and Stanley, 1998). Sequences necessary for vector transmission have been located in the central part of the protein (Liu et al., 2001; Qin, Ward, and Lazarowitz, 1998; Unseld et al., 2001). The absence or inactivation of CP generally results in reduced levels of viral ssDNA without reduction in the level of dsDNA, as observed in plants and protoplasts infected with CP mutants (Briddon et al., 1989; Brough et al., 1988). Because geminiviruses replicate in the nucleus of the infected host cells, following their inoculation into cytoplasm by the vector, the virus need to be transported into the nucleus for replication.

To be recognized by host receptors, these virus associated proteins must contain nuclear localization signals (NLS) (Gafni and Epel, 2002). Such signals have been determined for both monopartite and bipartite geminivirus and are mainly located in the N-terminal region of the CP. For ACMV, two other domains containing NLS, which are located in the central (100-127 amino acids) and C-terminal (201-258 amino acids) regions were also determined (Unseld et al., 2001). The CP of geminivirus also participates in exporting the viral genome from the nucleus to the cytoplasm. In this case the nuclear export signals (NES) is required for recognition by a host receptor. A NES signal located in the C-terminal half of the TYLCV CP has been identified (Rhee et al., 2000).

1.4.6 Pre-coat protein (AV2/V2)

In contrast to New World begomoviruses, Old World begomoviruses have AV2/V2 and this is involved in the movement of monopartite viruses but its function for bipartite begomoviruses is not fully understood (Rothenstein et al., 2006; Padidam, Beachy, and Fauquet (1996) showed by mutation analysis that AV2 is also involved in the movement of bipartite geminiviruses. Rothenstein, Krenz, and Selchow (2007) analyzed viral intercellular transport in further detail. GFP was fused to AV2 and expressed from replicating viruses or from plasmids, AV2: GFP became associated with the cell periphery in punctate spots, formed cytoplasmic as well as nuclear inclusion bodies, the later as conspicuous paired globules. Upon particle bombardment of expression plasmids, AV2: GFP was transported into neighboring
cells of epidermal tissues showing that the intercellular transport of the AV2 protein is not restricted to the phloem.

The absence of this gene in begomoviruses from the New World has been believed to consistent with the fact that all begomoviruses from the New World are bipartite in nature and suggest an evolutionary divergence of these viruses from begomoviruses of the Old World. But the analysis of the *Corchorus yellow vein virus* (CoYVV), a bipartite begomovirus infecting Jute mallow (*Corchorus capsularis, Tilliaceae*) in Vietnam showed the absence of AV2 ORF. Evidence is provided that CoYVV is probably indigenous to the region and may be the remnant of a previous population of New World begomoviruses in the Old World. A recent report shown that the V2 (a homolog of AV2) of a monopartite begomovirus is involved in overcoming host defenses mediated by post-transcriptional gene silencing as well as in movement. V2 targets a step in the RNA silencing pathway which is subsequent to the Dicer-mediated cleavage of dsRNA (Zrachya et al., 2006; Rojas et al., 2001).

Rojas et al. (2001) investigated the properties of proteins (CP, AV1, and C4) potentially involved with movement of the monopartite begomovirus. The TYLCV AV1 localized around the nucleus, at the cell periphery and co-localized with the endoplasmic reticulum. Thus, the AV1 may be an analog of the bipartite begomovirus BC1 that has evolved to mediate TYLCV movement within phloem tissue. Recently, Mubin et al. (2007) showed that ToLCNDV AV2 is an effective target through RNAi to engineer broad spectrum resistance against begomoviruses.

### 1.4.7 Nuclear shuttle protein (NSP) and movement protein (MP)

For the propagation of infection in the host, viruses encode movement proteins that direct the viral genome to the cortical cytoplasm and across the barrier of the cell wall. For bipartite begomoviruses, this process requires two proteins, encoded by the component designated DNA B: the nuclear shuttle protein (NSP) and the cell-to-cell movement protein (MP). These proteins act cooperatively to move the viral DNA from its site of replication in the nucleus to the cytoplasm and into adjacent plant cells (Sanderfoot and Lazarowitz, 1996).
The geminivirus MP and NSP recognize DNA in a form- and size-specific manner (Hehnle, Wege, and Jeske, 2004; Rojas et al., 1998). The NSP shuttles newly replicated viral DNA between the nucleus and the cytoplasm (Ward and Lazarowitz, 1999). MP traps these NSP-viral DNA complexes in the cytoplasm and redirects them to and across the plant cell wall (Noueiry, Lucas, and Gilbertson, 1994; Ward and Lazarowitz, 1999). In adjacent cells, the NSP-viral DNA complexes are released, and NSP targets the viral DNA to the nucleus to initiate new rounds of infection. This process of movement is highly regulated. Recently, an acetyltransferase has been isolated from *Arabidopsis thaliana* that interacts with the NSP encoded by the bipartite begomovirus CaLCuV (McGarry et al., 2003). This protein regulates the nuclear export of the viral DNA and is highly conserved in plants.

For some bipartite begomoviruses, MP is the major symptom determinant, and the expression of this protein induces disease-like symptoms (Brough et al., 1988; Duan et al., 1997a; Etessami et al., 1988; Hou et al., 2000; Ingham, Pascal, and Lazarowitz, 1995; Pascal et al., 1993). The NSP of *Bean dwarf mosaic virus* (BDMV) is reported to be an avirulence determinant inducing a hypersensitive response (HR) in *Phaseolus vulgaris* (Garrido-Ramirez et al., 2000). This established a key role for the NSP and MP gene products in systemic infection, as is the case with other bipartite begomoviruses. (Hussain et al., 2005) have analyzed the role of the MP and NSP of ToLCNDV in pathogenicity. They described the NSP of ToLCNDV is a symptom determinant and an avirulence determinant that is the target of host defense responses in tobacco and tomato leading to a HR.

1.5 **Satellites associated with geminiviruses**

Satellites are mostly associated with RNA viruses. Satellite viruses code for their own coat protein, whereas the RNA satellites use the coat protein of the helper virus for encapsidation. These molecules depend on a helper virus for replication and are dispensable for the replication of the helper virus. These molecules lack sequence homology to the helper virus’ genome (Murant and Mayo, 1982). The majority of satellites interferes with the replication of their helper viruses and attenuates symptoms. A small number of satellites, however, are known that exacerbate
symptoms or produce novel symptoms which are not usually associated with the helper virus infection (Collmer and Howell, 1992).

1.5.1 Betasatellite

For the first time for begomoviruses, *Tomato leaf curl virus* (ToLCV), originating from Australia, was shown to be associated with a single-stranded DNA satellite molecule (Dry et al., 1997). The ToLCV satellite (ToLCV-sat) is approximately 682 nt in length and sequence unrelated to ToLCV and it depends on ToLCV for replication and encapsidation. It has no discernable effects on viral replication or symptoms caused by ToLCV.

In contrast to the ToLCV-sat, betasatellites affect the replication of their helper begomoviruses and are required for the induction of disease symptoms in some host plants (Saunders et al., 2000; Briddon et al., 2001). The betasatellite associated with *Eupatorium yellow vein virus* (EupYVV) has been linked to disease symptoms described about 1250 years ago (Saunders et al., 2003). Full length clones of monopartite begomoviruses, *Ageratum yellow vein virus* (AYVV) from Singapore and *Cotton leaf curl Multan virus* (CLCuMV) from Pakistan, although infectious, were unable to induce typical symptoms of yellow vein in *Ageratum conyzoides* and leaf curl in cotton, respectively, and novel molecules, named betasatellites, were shown to be associated with both viruses, and to be essential for induction of characteristic symptoms in *Ageratum* and cotton (Saunders et al., 2000; Briddon et al., 2001).

Analysis of betasatellite molecules revealed that they are approximately half the size of their helper begomoviruses and except for a conserved hairpin structure and a TAATATTAC loop sequence, have little similarity to either DNA A or DNA B molecules of begomoviruses. Betasatellites require begomovirus DNA A for replication, encapsidation, insect transmission and movement in plants (Saunders et al., 2000; Briddon et al., 2001). Betasatellites have three structural features: a approx.115 bp highly conserved region, βC1 gene and a region rich in denine, (Saunders et al., 2000; Zhou et al., 2003). This gene has the capacity to encode a 13- to 14-kDa protein comprising 118 amino acids, although some betasatellites have additional N-terminal amino acids (Saunders et al., 2004; Zhou et al., 2003). The
Introduction and Review of Literature

Precise function of betasatellite and its βC1 protein in pathogenesis is not clear but recently it has been shown to be a pathogenicity determinant and suppressor of RNA silencing (Cui et al., 2005; Qazi et al., 2007). It also induced abnormal cell division in *N. benthamiana* (Cui et al., 2004). Betasatellites do not contain the iterons of their helper begomoviruses, although betasatellite clearly must possess sequences that are recognized by the begomovirus-encoded Rep in order to allow transreplication of the betasatellite (Saunders et al., 2000).

The DNA A components of the bipartite begomoviruses *African cassava mosaic virus* (ACMV) and *Indian cassava mosaic virus* (ICMV) are able to transreplicate Ageratum yellow vein betasatellite (AYVVB) betasatellite in *N. benthamiana*, but were unable to functionally interact with AYVVB to produce a symptomatic systemic infection (Saunders et al., 2002), indicating that there is less replication specificity for DNA A mediated replication and that many species of begomovirus could potentially replicate many kinds of betasatellites.

It is important to know how these molecules are recognized by the DNA A-encoded Rep proteins and how they control DNA A-mediated replication (Saunders et al., 2002). (Briddon et al., 2003) obtained 26 additional betasatellite molecules, associated with diverse plant species obtained from different geographical locations, were cloned and sequenced. These molecules were shown to be widespread in the Old World, where monopartite begomoviruses are known to occur. The SCR contains a potential hairpin structure with the loop sequence TAAGTATTAC; similar to the origins of replication of geminiviruses and nanoviruses. Two major groups of betasatellites were resolved by phylogenetic analyses. One group originated from hosts within the Malvaceae and the second from a more diverse group of plants within the Solanaceae and Compositae. Within the two clusters, betasatellites showed relatedness based both on host and geographic origin. These findings strongly support coadaptation of betasatellite molecules with their respective helper begomoviruses.

All the reported betasatellites (Mansoor et al., 2003b) or defective betasatellites (half size of wild type betasatellite) (Briddon et al., 2003) contain the A-rich region, the a-rich region may play biological role in betasatellites (Tao, 2004). A-rich region is not required for trans-replication of betasatellite and not related with
encapsulation also. However, the A-rich region deleted mutant caused milder symptom (Tao and Zhou, 2004). The begomovirus accumulates to normal levels in *Ageratum* in the presence of betasatellite suggesting that the satellite functions either by facilitating the replication or movement of the begomovirus or by suppressing a host defense mechanism such as gene silencing. Recently it has been shown that a betasatellite can override the AC4 pathogenicity phenotype of ToLCV and it can complement the function of DNA B (Saeed, 2008). Despite its importance to the disease phenotype, there is still no information available concerning even the most fundamental properties of the satellite. The satellite encodes at least one protein that plays an important role in the pathogenicity of the begomovirus-satellite disease complex (Saunders et al., 2004). βC1 protein may also play a role in developmental regulation by interfering with miRNA pathways.

In *Arabidopsis*, these pathways are affected by the DICER-like proteins (DCL1, DCL2, and DCL3) that are nuclear localized and are required for miRNA and siRNA biogenesis. Thus, βC1 protein may affect the activity of the DICER-like proteins in plants during nuclear activities that function in silencing suppression. The other possibilities are that βC1 protein could down-regulate transcription of a host protein that acts in the PTGS pathway in the cytoplasm or that βC1 protein could activate transcription of a host PTGS inhibitor. (Cui et al., 2005).

### 1.5.2 Alphasatellite

Begomovirus-betasatellite complex are associated with alphasatellites. These satellite-like molecules encode a single product which shows similarity to the replication associated protein (Rep; a rolling-circle replication initiator protein) of nanoviruses; another family of plant infecting single-stranded DNA viruses. These molecules can replicate autonomously but require a helper begomovirus for encapsidation and insect transmission. Alphasatellites appear to have no role in the disease process, being dispensable both for infectivity and symptom induction in host plants (Briddon et al., 2004).
Figure 1.5 Genetic structure of betasatellites. Shown are the βC1 gene (βC1), the adenine rich sequence (A-rich) and the satellite conserved region (SCR) which is highly conserved among all the betasatellites. The hairpin structure, containing the conserved nonanucleotide sequence (TAATATTAC) is shown at position ‘O’ in the SCR.

Figure 1.6 Genetic structures of alphasatellite. Shown are the Rep (replication associated protein); the adenine rich sequence (A-rich). The hairpin structure, containing the conserved nonanucleotide sequence (TAGTATTAC) is shown at positon ‘O’.
1.6 Transmission of geminiviruses

Geminiviruses depend on insect vectors for their transmission but there are some viruses which are reported to be vegetatively transmitted through infected stocks, as geminiviruses associated with cassava mosaic disease. Begomoviruses are transmitted by the whitefly, *Bemisia tabaci* (Gennadius). *B. tabaci* was first described in 1889 (Gennadius, 1889). All begomoviruses are transmitted *B. tabaci*; the efficiency is different for different viruses and suggests adaptation of a virus to a whitefly biotype. Over 100 begomoviruses are transmitted by at least two biotypes of *B. tabaci* to more than 20 cultivated species of socioeconomic importance. Begomovirus transmission by *B. tabaci* is circulative and non-propagative (Markham et al., 1994; Gray and Gildow, 2003). As the feeding time increases the amount of virus acquired by the whitefly increases. Studies of the transmission of *Tomato yellow leaf curl virus* (TYLCV), a monopartite begomovirus, showed that whitefly feeding periods of 4hrs or longer were necessary to achieve TYLCV transmission rates near to 90% (Zeidan and Czosnek, 1991). Hunter et al. (1998) proposed a model for the movement of begomoviruses in the whitefly vector. Virus particles are ingested along with plant fluids into the whitefly esophagus and foregut. As food enters the filter chamber excess water is shunted to the ileum of the hindgut. Thus nutrients and begomoviruses are concentrated in the filter chamber. Begomovirus particles adsorb to specific sites on the alimentary membrane or to sites along the anterior region of the midgut. Begomovirus particles move out of these tissues into the hemolymph, eventually invading the salivary glands. There is no evidence for viral replication in the insect vectors. The acquisition of virus by whitefly has a cytoplasmic effect on the insect and decreases the average life span of the insect (Czosnek, 2001).

There are many species of sucking insects which can acquire viruses but a few are actually capable to transmit to the next host (Liu et al., 1997). There are a few species of leaf hoppers (12-18) and one species of treehopper which can transmit geminiviruses. It is actually coat protein which determines the specificity of geminivirus transmission by an insect vector. Some viruses have lost their ability to be transmitted by insect vector like *Abutilon mosaic virus* and *Honey suckle yellow*
vein virus. This lack of insect transmission was attributed to a defective in the CP (Kleinow et al., 2008).

Briddon et al. (1990) replaced the CP gene of ACMV with that of BCTV and the resulting virus was infectious and leafhopper transmissible. There is more variation in the coat protein of leafhopper transmitted geminiviruses and suggests a correlation with leafhopper species. Curtoviruses are transmitted by the beetleafhopper, *Circulifer tenellus* (Baker). Soto et al. (2005) described the N-terminus of *Beet mild curly top virus* (BMCTV) involved in virus movement through the leafhopper, possibly receptor-mediated endocytosis in the gut or salivary glands. Another region of the BCTV CP, residues 179–191, was postulated to be an adaptation for leafhopper transmission (Böttcher et al., 2004).

### 1.7 Geminivirus-host interactions

Geminiviruses depend for their DNA replication and transcription on host machinery. The identification of host factors which interact with viral proteins is of much interest especially for those people working on engineering resistance against these viruses. However, the host factors required for viral processes and the mechanism whereby geminiviruses become established in plant nuclei are very well known. There have a number of recent studies which have begun to provide an insight into how geminiviruses interact with and modify their host during infection. AC1 is a multifunctional protein, which interacts with host cell factors to interfere with control of cell cycle and DNA replication in the infected cells (Gutierrez, 2000).

The Rep proteins of geminiviruses are closely related and show substantial sequence conservation. This protein is solely responsible for the RCR of geminiviruses. Most of the host factors crucial for RCR initiation and control and viral pathogenesis have been found to interact with the virus encoded AC1 protein. The AC1 proteins of many geminiviruses have been found to interact with plant retinoblastoma-like proteins (pRBR) to alter the cell cycle programs of infected non-dividing cells (Ach et al., 1997). The majority of the geminiviruses infect terminally differentiated cells that have exited the cell division cycle and contain very small amounts of DNA replication enzymes (Castillo et al., 2003). Due to this interaction
infected G0 cells are activated to partial S phase (Egelkrout et al., 2002) thus ensuring viral replication processes. At least one important host replication factor, proliferating cell nuclear antigen (PCNA), an accessory to DNA polymerase δ, has been found to be induced by this mechanism following infection by TGMV (Kong and Hanley-Bowdoin, 2002). One type of virus-host interaction that is well established and widespread is the modulation of viral protein function by posttranslational modification systems such as phosphorylation, glycosylation, ubiquitination, and sumoylation (Castillo et al., 2004).

AC1 and AC3 of TGMV interact with each other and themselves and independently interact with the host protein pRBR, the plant retinoblastoma homolog and proliferating cell nuclear antigen (PCNA) (Hanley-Bowdoin, Settlage, and Robertson, 2004) but the AC3-RBR interaction is not required for viral replication in cycling cells (Settlage, See, and Hanley-Bowdoin, 2005).

Wheat dwarf virus (WDV) RepA interacts with two wheat (Triticum aestivum) proteins, GRAB1 and GRAB2 which are involved in diverse developmental processes (Xie et al., 1999). Recently, ToLCV and TGMV AC3 were shown to interact with a transcription factor in the NAC family (Selth et al., 2004). AC2 is also a multifunctional protein and is only present in begomoviruses. AC2 also has been shown to regulate transcription of host genes (Trinks et al., 2005). AC2 was shown to suppress silencing indirectly by activating the expression of a cellular protein that may function as an endogenous negative regulator of the system. Proteins from two different geminiviruses, AC2 from TGMV and C2 from BCTV are are homolog of begomovirus AC2 and are pathogenicity determinants (Sunter, Sunter, and Bisaro, 2001), which is attributable to their ability to inactivate SNF1-related kinase (Hao et al., 2003; Sunter, Sunter, and Bisaro, 2001)) AC2 and C2 also interact with and inactivate adenosine kinase (ADK), which phosphorylates adenosine to produce 5-AMP (Wang et al., 2005). Because AMP can stimulate SNF1 activity, the inactivation of SNF1 and ADK by AC2/ C2 may represent a dual mechanism to counter SNF1-mediated antiviral responses. ADK activity is required to support RNA silencing, and indicate that the geminivirus proteins suppress silencing by a novel mechanism that involves ADK inhibition. The expression of AC4/C4 protein in transgenic N.
*benthamiana* produced virus-like symptoms and further confirmed its role in symptom development (Latham et al., 1997).

Chellappan, Vanitharani, and Fauquet (2005) provided direct evidence that AC4 is a unique virus-encoded, PTGS suppressor protein that binds to and presumably inactivates mature miRNAs and thus blocks the normal miRNA-mediated regulation of target mRNAs, resulting in developmental defects in *Arabidopsis*. In the yeast two hybrid system it was shown recently that *Beet curly top virus* (BCTV) C4 interacts with two members of the shaggy-related protein kinase family (AtSKη and AtSKζ) and a putative leucine-rich repeat receptor-like kinase (LRR-RLK) involved in autophosphorylation. This interaction shows that BCTV C4 interacts with the brassinosteroid signalling pathway (Piroux et al., 2007).

As far as coat protein interaction with host factors concerned, recently the CP protein of ToLCV and a protein closely related to a family of plant reversibly glycosylated peptides, designated SIUPTG1 were found interacting in yeast and in vitro. Expression of SIUPTG1 in a transient ToLCV replication assay increased the accumulation of viral DNA. This data provided new insights into the role of CP in ToLCV infection and reveal another host pathway that geminiviruses may manipulate to achieve an efficient infection (Selth et al., 2006).

AV2/V2 is found only in Old World begomoviruses and this is involved in the movement of monopartite viruses but its function for bipartite begomoviruses is not fully understood (Rothenstein et al., 2006). A recent report has shown that the V2 (a homolog of AV2) of a monopartite begomovirus is involved in overcoming host defenses mediated by post-transcriptional gene silencing as well as in movement. It was shown that AV2 interacts directly with SISGS3, the tomato homolog of the Arabidopsis SGS3 protein (AtSGS3), which is known to be involved in the RNA silencing pathway (Glick et al., 2007).

For the propagation of infection in the host, viruses encode movement proteins that direct the viral genome to the cortical cytoplasm and across the barrier of the cell wall. NSP and the MP perform this function in bipartite begomoviruses. Recently, an
acetyl transferase has been isolated from *Arabidopsis thaliana* that interacts with the NSP encoded by the bipartite begomovirus CaLCuV (McGarry et al., 2003).

### 1.8 Movement of geminiviruses

Plant viruses have to be more adaptive than animal viruses in their spread from cell to cell as plant cell is surrounded by stiff cell wall. These viruses have to move cell to cell and ultimately enter the phloem to infect the whole plant. Viruses do this job by encoding a unique class of proteins termed “movement proteins”. Movement proteins determine the host range and disease potential of the virus. The first method of viral spread was most likely associated with the cell division which could be the dated with the earlier forms of multi cellular plants. All cell types are interconnected through plasmodesmata but there is always a limit of size of molecule which can pass through these channels. The movement proteins encoded by these viruses have been shown to be nucleic acid binding proteins and some of these can increase plasmodesmatal size exclusion limit (SEL) to enable the virus or viral nucleo proteins complexes to move through the plasmodesmata. The establishment of a virus infection depends upon the spread of the virus throughout the host plant. The movement of the virus in the plant occurs at two different levels. Firstly short distance cell-to-cell movement, secondly long distance movement in which virus is moved through vascular system to the all parts of the plant (Lazarowitz, 1992).

There are different mechanisms and ORFs involved in the movement of geminiviruses for different genera. CP mediated systemic movement has been reported for all genera, showing that geminiviruses move as particles in long distance movement. CP is a multifunctional protein as it functions in systemic infection and insect transmission. Consistent with its role in encapsidation the CP is nuclear localized, via N-terminal NLS and an interaction with importin family alpha. In begomoviruses an additional ORF precoat protein is involve in the cell to cell movement of viruses (Padidam, Beachy, and Fauquet, 1996). TYLCV and EACMV V2-GFP fusions showed a perinuclear distribution, co-localization with endoplasmic reticulum and accumulated at the cell periphery (Zrachya et al., 2006). Micro injection experiments demonstrated the AV2 enhanced the CP trafficking through the
mesophyll PD. These results show that AV2 mediates viral DNA export from the nucleus to the PD.

In bipartite begomoviruses the DNA B, specialized for the movement, encodes two movement proteins, MP and NSP. Genetic, protoplast, transgenic plants and GFP tagged viruses studies established that these proteins are not required for replication or encapsidation, but are essential for efficient movement. NSP and MP coordinate the movement of viral DNA across the nuclear and plasmodesmatal boundaries, respectively. Sanderfoot and Lazarowitz (1995) proposed for bipartite begomoviruses that the NSP of SqLCV binds the replicated viral DNA in the nucleus, exports this to cytoplasm where MP specifically binds the NSP-DNA complexes and directs this to the cell periphery, where they are transported into the adjoining cells and to the phloem tissue. The localization of NSP is consistent with this function of NSP. The NSP is localized to nucleus and nucleolus of infected cells. It has two NLSs in N-terminus. NSP binds DNA on the basis of size and form depicting the mechanism of begomovirus genome size limitation. Sub cellular localization studies showed that MP is targeted to the perinuclear area, the cell periphery and punctuate bodies. Microinjection studies have shown that MP can move cell to cell and mediate cell to cell movement of ssDNA and ds DNA. The requirement of both proteins for intercellular movement was also demonstrated for BDMV, where mutation of the NSP and MP restricted the cell-to-cell movement of viral DNA (Sudarshana et al., 1998). The MP of BDMV increases the SEL of plasmodesmata and the protein mediates viral DNA transport from cell-to-cell (Noueiry, Lucas, and Gilbertson, 1994; Rojas et al., 1998). In contrast, the MP of SqLCV does not bind DNA but is present in virus-induced tubules that cross the walls of meristematic phloem tissues (Ward et al., 1997).

1.9 Impact of begomoviruses

1.9.1 Worldwide epidemics associated with begomoviruses

Epidemics of begomoviruses have increased in number, prevalence and distribution during the past two decades (Brown and Bird, 1992; Harrison, 1985; Otim-Nape, Thresh, and Shaw, 1997). They usually have a devastating impact on
agricultural and horticultural crops. Several factors are reported to be responsible for outbreaks of begomovirus diseases, including the emergence of new viruses and whitefly biotypes, increased use of pesticides to control vector pests, monoculture, and the use of cultivars that are not tolerant or resistant to geminiviruses. Before the 1980s, begomoviruses in the New World were primarily a problem for legume production (Polston et al., 1997). Begomoviruses have now been reported to continuously spread through the Western Hemisphere.

The most affected crops include cucurbits, tomato, and cotton in the USA, Caribbean, Mexico, Central America, Brazil, and Venezuela, which have suffered from high incidences of begomoviruses with devastating economic consequences. In addition, crops such as potato, melons, and beans are also affected. In 1990s geminiviruses destroyed up to 95% of the tomato harvest in the Dominican Republic, and in 1991-92 growing season in Florida, they caused $140 million in damage to the tomato crop (Fauquet and Stanley, 2003).

In Sudan, epidemics of cotton leaf curl disease (CLCuD) were reported in 1940s and 1950s and continue to threaten the cultivation of cotton, the most economically important crop of the country (Idris and Brown, 2000). In Pakistan, CLCuD has acquired epidemic proportions (Mansoor, 1999). The epidemic of CLCuD coincided with massive infestations of B. tabaci and the widespread cultivation of extremely virus susceptible cotton varieties in Pakistan (Harrison and Robinson, 1999). This disorder of cotton was first noted in Pakistan in the late 1960s. It remained a minor sporadic problem for the following 20 years. In 1988 a small plot of a newly released cotton variety (S12), grown at Moza Khokran near Multan, was severely affected with CLCuD. This signaled the beginning of the epidemic, which spread to all cotton growing areas of Pakistan (Hussain and Mahmood, 1988).

The CLCuD epidemic in Pakistan that started with the natural host range limited to a few plant species in the family Malvaceae has been found to expand as there are evidence of mobilization of CLCuD into new host plants like radish, watermelon, Croton bonplandianus (family Euphorbiaceae) and okra (Amin et al., 2002; Mansoor et al., 2000b). Over the past few years, the introduction of CLCuD resistant/tolerant cotton varieties in Pakistan produced by conventional
breeding/selection methods, showed promise in the management of the disease. However, recently resistance has broken down, as previously resistant varieties are now showing symptoms of CLCuD in Pakistan (Mansoor et al., 2003c).

In the Middle East and Mediterranean countries, production of vegetables especially tomato was hampered by epidemics of tomato leaf curl disease (ToLCVD) (Czosnek and Laterrot, 1997). The epidemics were usually associated with the introduction of a virus or vector strain into new agroecological conditions. For example, in Spain, an epidemic of ToLCVD occurred with the introduction of Tomato yellow leaf curl virus-Israil (TYLCV-Is) by the B-biotype, which spread more rapidly than the endemic tomato yellow leaf curl Sardinia virus (TYLCSV). The epidemic occurred rapidly because the tomato cultivars grown were susceptible to TYLCSV in Spain. Recently, a new ToLCD epidemic was also reported in India. The increasingly wide distribution of higher population levels of whitefly vectors in agroecosystems are directly implicated in the development of new geminiviruses epedemics throughout the world (Polston et al., 1997).

1.9.2 Impact of begomoviruses associated diseases in Pakistan

Virus diseases, in general, are a major factor limiting crop productivity in Pakistan. Among different viruses, begomoviruses are a major threat found to be associated with a number of economically important crops (Sanz et al., 2000) ranging from the country’s cash crops like cotton to vegetables like tomato, chillies, radish, cucurbits as well as ornamental plants like hibiscus and leguminous plants like mungbean (Mansoor et al., 2000a; Mansoor et al., 2003a).

Cotton leaf curl disease (CLCuD) The disease is, at this time, endemic throughout Pakistan and epidemic in Western India. CLCuD has been shown to be associated with several viruses of the genus begomovirus that were collectively referred to as “cotton leaf curl virus” (Briddon and Markham, 2000; Mansoor, 1993; Zhou et al., 1998). Cotton has a higher cost of production than any other crop with 66 per cent of the total cost of production per hectare incurred on chemicals for plant protection (International Cotton Advisory Committee, 1994). During the 5 years between 1992 and 1997 the cost to the Pakistan economy due to CLCuD has been
Introduction and Review of Literature

estimated at US$5 billion. In Pakistan yield losses due to CLCuD as high as 100% in cotton fields (Mansoor, 1999).

Much economic loss of tomato crops in Pakistan is attributed to ToLCD (Mansoor et al., 2005). ToLCD disease is most important constraint for tomato production in Pakistan, with losses estimated to 30-40% during spring while in autumn it had become uneconomical due to severe yield losses (Mansoor, Khan, and Saeed, 1997).

1.10 Recombination among begomoviruses

Geminiviruses are an increasing threat to crop plants worldwide, especially in tropical and subtropical countries (Moffat, 1999). One reason for this is the recombinogenic nature of geminiviruses (Zhou, Robinson, and Harrison, 1998). Footprints of recombination were found upon sequence comparisons of several geminiviruses (Padidam, Sawyer, and Fauquet, 1999). Sequence comparisons between the different genera of family Geminiviridae have led to the suggestion that curtoviruses have evolved from a recombination of an ancient mastrevirus and begomovirus (Stanley et al., 1986). All is evidence suggests that recombination might be a major driving force for the evolution of geminiviruses and their ability to break resistance in crop plants (Harrison and Robinson, 1999).

It has been speculated that recombination between geminiviruses has been a major contributing factor behind the recent emergence of a number of devastating crop diseases worldwide (Amin et al., 2006; Padidam, Sawyer, and Fauquet, 1999). Survival of the fittest is the rule of the nature. The long-term survival and establishment of a recombinant virus genotype in nature depends strongly on it having a selective advantage over the overwhelming populations of its parental genotypes (Schnippenkoetter et al., 2001).

Cotton leaf curl disease was originally a major problem in central Pakistan but is now causing extensive damage in India. In the same region, new diseases are emerging in crops such as tomato, tobacco, chilies and papaya. The presence of such a diverse population of begomoviruses in a single region, coupled with the propensity of these viruses to exchange genetic material by recombination (Roberts and Stanley,
1994; Saunders, Bedford, and Stanley, 2001; Saunders, Bedford, and Stanley, 2002) increases the probability of new virus diseases emerging to cause epidemics in previously unaffected crops.

In experiments, recombination occurred quickly if two handicapped virus constructs were co-inoculated (Evans and Jeske, 1993). Recombination was also detected between a mutated virus coat protein (ACP) gene of ACMV and its homologous transgene (CP) in *Nicotiana benthamiana* plants (Frischmuth and Stanley, 1998). This phenomenon is called forced recombination. This process has also been reported to occur between two distinct isolates of *Maize streak virus* (MSV). The two most virulent recombinants were leafhopper transmitted to a range of differentially MSV-resistant maize, wheat and barley genotypes and both were found to have unique biological properties (Schnippenkoetter et al., 2001).

Recent evidence revealed the presence of a naturally occurring recombinant DNA-A of a typical bipartite begomovirus, *Tomato chlorotic mottle virus* (ToCMV) that does not require the cognate DNA B to infect *N. benthamiana* systemically. ToCMV DNA A did not infect tomato plant however it systemically infected *N. benthamiana*, induced symptoms of mottling and accumulated viral DNA in the apical leaves in the absence of a cognate DNA B (Galvao et al., 2003).

Recombination, thus, is a powerful factor in the evolution of begomoviruses, not only in the long term but also in short and medium term (Zhou et al., 1997). Recombination among begomoviruses (Zhou et al., 1997; Harrison and Robinson, 1999; Padidam, Sawyer, and Fauquet, 1999; Sanz et al., 1999) can thus provide additional sources of variation with unpredictable effects on virus pathogenicity.

1.11 **Favored features for recombination among geminiviruses**

1.11.1 **Bipartite genome**

Geminiviruses having bipartite genomes are in abundance, and this feature may facilitate viral evolution through pseudo recombination or recombination. The pseudorecombinants resulting from mixing up DNA A and DNA B components of the two geminiviruses BDMV and *Tomato mottle virus* (ToMoV), were found to be
infectious in *N. benthamiana* although the symptoms were not so severe and the level of DNA B was also reduced.

Sequence analysis of the DNA B component of the more pathogenic pseudo recombinant revealed that BDMV DNA B common region was replaced with the ToMoV DNA A common region. This recombinant DNA B component, which contained the ToMoV origin of replication, was the predominant DNA B component associated with the more pathogenic pseudorecombinant. These results demonstrate that recombination events can take place between distinct bipartite geminiviruses and establish that the bipartite genome also facilitate viral evolution through pseudorecombination and intermolecular recombination (Hou and Gilbertson, 1996).

### 1.11.2 Multiple/mixed infections

Mixed infections that favour recombination and pseudorecombination among viruses have also been observed in begomoviruses (Lazarowitz, 1991; Harrison et al., 1997; Harrison and Robinson, 1999; Roye et al., 1999). Recombination, following multiple infections, could also explain the network of relationships among many of the begomoviruses found in the Indian subcontinent, and their evolutionary divergence, as a group, from begomoviruses causing similar diseases in other geographical regions (Sanz et al., 2000). There are many examples of begomoviruses with hosts in common. Numerous opportunities for multiple infections and recombination may exist where these viruses occur in same geographical areas.

Mixed infection can contribute towards disease severity as different viruses might display synergistic interactions. As in case of CMD, the mixed-infected samples always showed extremely severe CMD symptoms, suggesting a synergistic interaction between virus strains/species (Pita et al., 2001). Recent begomovirus epidemics reflect favorable conjunctions of plant, vector, and viral (such as the emergence of a novel recombinant virus) factors. Such epidemics typically result in co-infection of plants with different begomoviruses, leading to the appearance of further variants, especially recombinants (Harrison and Robinson, 1999). Such multiple infections events may thus be contributing to the emergence of new forms of
begomoviruses, which are continuously being recorded such as Central America, southern United States and Pakistan (Brown and Bird, 1992; Mansoor et al., 2003c).

1.11.3 Acquisition of satellite DNA components

Satellite DNA molecules have been found to be associated with many begomovirus diseases and sufficiently contribute towards disease epidemics (Mansoor et al., 2003b). Genomic variation resulting from mutation in geminiviruses is amplified by acquisition of extra DNA components (i.e. satellite DNAs), pseudo-recombination and recombination, both intraspecific and interspecific (Harrison and Robinson, 1999). Evidences suggest that betasatellites may have a relaxed relationship with begomoviruses rather than the specific interaction characteristic of DNA A and DNA B components. This satellite component can be recruited by diverse begomoviruses, resulting in multiple virus species supporting a single satellite. The recruitment by other begomoviruses is likely to occur through co-infection of alternate host.

It has been reported that CLCuMB can interact with and is transreplicated by different monopartite begomoviruses inducing symptoms of the CLCuD. Recombination events can result in more pathogenic satellites as a recombinant betasatellite has found to be associated with the resistance breakdown in cotton cultivars against CLCuD. Moreover relaxed specificity of betasatellite for begomovirus, mobilization from one host to other with increased pathogenecity of these molecules, have also been reported (Mansoor et al., 2003a; Amin et al., 2006).

1.11.4 Insect vector

One reason for the appearance of epidemics is the spread of the vectors of geminiviruses (Rybicki and Pietersen, 1999). Geminiviruses have evolved a highly dependent relationship between their host plant and the whitefly vector. Whitefly has the potential to colonize a wide range of dicotyledonous species. Recent studies indicate that there are numerous populations of B. tabaci that vary somewhat in their capacity to develop high population densities and cause direct feeding damage, in the extent of their host ranges, and in the efficacy with which they can transmit geminiviruses (Bedford et al., 1994; Brown et al., 1995).
The establishment of the B biotype in cotton-vegetable agro-ecosystems is driving force behind the emergence of geminiviruses in cotton-vegetable agro-ecosystems (Brown et al., 1995). The increasingly wide distribution of higher population levels of whitefly vectors in agro-ecosystems are directly implicated in the development of new geminiviruses epidemics throughout the world. An important consequence of all such epidemics is a much increased incidence of multiple infections of viruses, which not only may increase disease severity, but also increase the chances of the emergence of new recombinant viruses (Harrison and Robinson, 1999).

1.1.2 Replication of geminiviruses

Geminivirus DNA replication follows a rolling circle mechanism. The rolling circle replication (RCR) of geminivirus can be divided into two phases (Gutierrez, 2000). First, conversion of viral ssDNA into dsDNA forms on entering the nucleus of initially infected cells. This step of synthesis of viral minus strand is carried out by cellular enzymes. Second, Rolling circle phase to replicate viral ssDNA on dsDNA templates (fig. 1.7). This step requires the participation of Rep. Rep is the only viral protein absolutely required for RCR, as it is responsible for initiating DNA replication. (Laufs et al., 1995) described in detail the role of Rep in initiation and termination of RCR of geminiviruses.

Recently an additional method of replication of geminiviruses and their satellites has been identified (Alberter, Rezaian, and Jeske, 2004; Jeske, Lutgemeier, and Preiss, 2001; Preiss and Jeske, 2003). The model proposed for this, recombination-dependent replication (RDR), was based on analysis of replication intermediates of AbMV, BCTV, TGMV, ACMV, ToLCV and one betasatellite, using two dimensional gel electrophoresis and electron microscopy.

Apart from previously identified RCR intermediates (Saunders, Lucy, and Stanley, 1991) a range of intermediates suggested an additional pathway. This is analogous to the RDR pathway of T4 bacteriophage (Kreuzer, 2000) that has also been named the “join-copy” pathway (Jeske, Lutgemeier, and Preiss, 2001), “break-induced replication” and “bubble-migration synthesis” (Formosa and Alberts, 1986).
The RDR model has three steps (Kreuzer, 2000; Jeske, Lutgemeier, and Preiss, 2001). Processing of the broken dsDNA to produce the 3’ end ssDNA required for DNA strand invasion. Invasion of a homologous duplex by 3’ end ssDNA to form a structure known as the ‘displacement loop’ (D-loop or bubble loop). DNA strand invasion by the third end of ssDNA allows it to serve as potential primer for DNA replication. DNA heteroduplex extension (branch migration). At this step the protein directed branch migration occurs at the rear of the loop as DNA polymerase extends the leading strand product at the front of the loop. Because both reactions occur at a similar rate, the size of the loop is roughly unchanged. This type of RDR does not need topoisomerase, even when the circular DNA templates are supercoiled, and the two parent strands do not need to separate from each other (Kreuzer, 2000). RDR of geminiviruses apparently does not require participation of Rep in terms of its cognate virus recognition and nicking of ssDNA at the nonanucleotide sequence for initiation of replication. This possibility is also supported by a recent study (Lin et al., 2003) in which mutants of ToLCV and its sat-DNA molecule that were impaired in their ability to bind Rep in vitro, were still infectious to tomato.
Figure 1.7 Geminivirus DNA replication cycle and intercellular movement of viral DNA. Geminivirus DNA replication occurs in two stages. First, the ssDNA is converted into dsDNA with the participation of cellular factors. The dsDNA serves as template for viral gene expression. Secondly, the dsDNA initiates the rolling circle phase, with the participation of viral and cellular factors, to produce new ssDNA products. These can (i) re-enter the DNA replication pool, (ii) associate with CP or (iii) be transported outside the nucleus and to the neighbouring cell, most probably through plasmodesmata, with the help of viral MPs. This diagram is based on figure 2 in Hanley-Bowdoin (1999).
1.13 Natural plant defense against viruses

In their natural environments, plants are routinely challenged by parasites including viruses, bacteria, fungi, nematodes and insects, all of which have the molecular capability to manipulate host plants for their own sustenance. Plants have evolved an innate disease resistance involving a complex array of constitutively expressed resistance (R) genes to detect foreign invaders and defend themselves (Dangl, Dietrich, and Richberg, 1996; Holub and Cooper, 2004).

Physiological studies have been complemented by the molecular characterization of several R genes. Some of the downstream signalling components, including kinases, putative lipases, phosphatases and transcription factors have been identified. These local defence responses are amplified throughout the plant with the help of well known signal molecules such as salicylic and jasmonic acid, ethylene and MAP kinases (Lam, Kato, and Lawton, 2001; Dangl, Dietrich, and Richberg, 1996; Van Den Abeele et al., 2003).

Production of reactive oxygen species (ROS), including H$_2$O$_2$ is one of the earliest known responses in incompatible interactions between pathogens and plants. The ROS induce the accumulation of salicylic acid (SA) and trigger pathogenecity related (PR) protein expression (Epple et al., 2003; Pennell and Lamb, 1997). SA is an essential signaling hormone for activation of local and systemic defenses against pathogens in many plant species (Gruner, 2003; Pennell and Lamb, 1997; Lam, Kato, and Lawton, 2001). SA, however, is not the essential mobile signal transmitted through plants to initiate systemic defense (Vernooij et al., 1994). Recent identification of a lipid transfer protein implies the nature of the long distant signaling molecule in Arabidopsis may be lipid-derived.

Post-transcriptional gene silencing (PTGS) also known as RNA interference (RNAi) is an intrinsic plant defense mechanism can be efficiently triggered by double stranded (ds) RNA-producing transgenes and can provide high levels of virus resistance by specific targeting of cognate viral RNA. PTGS is a homology-dependent RNA degradation process that targets RNA exclusively in the cytoplasm.
1.14 Control of plant diseases

Plant diseases caused by infectious viruses, bacteria, fungi, phytoplasmas, and nematodes cause major problems in agricultural crops. These problems include reduced yields, lower product quality or shelf-life, decreased aesthetic or dietary value, and, sometimes, food and feed contaminated with toxic compounds. Control of a plant disease is essential for providing an enough supply of food, feed, and fiber. Growers currently spend large sums to attain partial control of pathogens that attack crops and other plants. Even then, crop and commodity losses because of diseases cost billions of dollars each year. Reducing such losses has long been a high main concern objective for agriculture. Knowledge and management of plant diseases of quarantine importance are essential, not only for protecting our domestic crops from distant disease, but also for maintaining and expanding export markets for plants and plant products.

Strategies for the control of plant diseases include sowing resistant crop varieties, changing cultural practices or storage conditions to those less encouraging for disease development, planting dates, employing biological controls, applying chemical pesticides, and using integrated disease management (combining two or more of the above approaches). The ability to develop any of these strategies depends first on identifying the pathogen causing the disease, then learning how to interrupt the disease cycle (Rahman, et al., 2001).

1.14.1 Genetically engineered resistance against viruses

Diseases caused by plant viruses are not easy to manage and their control mainly involves the use of insecticides to destroy insect vectors, the use of virus-free propagating materials, and the selection of plants with appropriate resistance genes. While insecticides can control vectors, often the virus has already been transmitted to the plant before the insect vector is killed. The use of resistant cultivars has been the most efficient means of control, however plant virus resistance genes are commonly unavailable and their introgression into some crops is not straight forward (Sharma, et al., 2000).
Genetic engineering has emerged as a “million dollar technology” to fight against viruses in comparison to conventional breeding. Various strategies have been engaged to develop genetically engineered resistance against plant viruses (Prins et al., 2008). Among them, there are pathogen derived resistance strategies and non-pathogen derived resistance strategies. Both of these approaches are stimulated certain amount of interest and also prove to be promising against a number of viruses (Sanford and Johnson, 1985).

1.14.2 Pathogen derived resistance

ssDNA viruses cause a broad range of disease in crop plants and contribute a major share in poverty and economic instability of this world. These viruses pose a threat to the quality of life and nutritional well being of the most of the population of the world. This situation makes the genetically engineered resistance against these viruses a prime objective for the well being of humanity. People have observed that in the phenomenon of cross protection, inoculation of a host plant with a milder or symptom less strain can protect the plant from infection by more severe strains of the same or very closely related viruses (Praveen, et al., 2004).

From these observations the concept of the pathogen-derived resistance arose which stated that certain key gene products of a pathogen present in the plant in a non-functional form, in excess, or at an inappropriate stage during viral replication cycle could disrupt infection by the invading pathogen. Pathogen-derived resistance is mediated either by the protein encoded by a transgene (protein-mediated) or by the transcript produced from the transgene (RNA-mediated) (Smith et al., 1994).

1.14.3 Protein-mediated pathogen-derived resistance

This type of pathogen-derived resistance employs the use of viral genes through the expression of their protein product. They include various viral genes like coat protein gene, (Beachy, 1990) sequences from the viral replicase gene (Nguyen et al., 1996), Rep protein gene (Noris et al., 1996; Brunetti et al., 1997) viral movement protein gene (Baulcombe, 1996) that have been used as a source of pathogen derived resistance in various plant species.
Geminivirus movement proteins are required for cell to cell and long distance movement of virus and have been used to engineer resistance against these viruses. Transgenic tobacco plants expressing the coat CP gene of *Tobacco mosaic virus* (TMV) were found to be more resistant to infection by TMV than the non-transgenic controls (Beachy, 1999). This initial demonstration of the feasibility of CP-mediated protection was followed by extensive research to develop genetically engineered CP-mediated plant virus resistance in various systems. The different viruses for which resistance has been demonstrated represent at least 13 different groups and include positive sense, negative sense, single- and double-stranded RNA viruses and at least one DNA virus (Nguyen, Ding, and Zaitlin, 1996). The level of protection conferred by CP genes in transgenic plants varies from immunity to delay and attenuation of symptoms. In some cases protection is broad and effective against several strains of the virus from which the CP gene is derived, or even against closely related virus species (Beachy, 1990; Lomonossoff, 1995). Despite extensive studies, the molecular mechanisms that govern CP-mediated resistance (CPMR) are not fully understood, and furthermore the mechanisms of CPMR are different in different viruses.

Transgenic plants expressing CP showed high resistance to challenge by virions, but not to inoculation with RNA or partially stripped virions; (ii) transgenic plants expressing TMV CP showed greater levels of CPMR against closely related viruses than to more distantly related viruses. Transgenic plants expressing mutant CPs affecting electrostatic interactions between the subunits showed modified CPMR according to their self-assembly capacity (Bendahmane and Gronenborn, 1997).

Native or altered virus derived genes might be used to interfere with various stages in the viral life cycle such as uncoating, translation, replication, cell-to-cell or long-distance movement, or vector-mediated transmission. Success has been reported with geminivirus transgene of modified and unmodified cp (Hong and Stanley, 1996; Kunik et al., 1994). Transgenic tomato plants expressing TYLCV CP were found resistant to the virus. In some studies resistance was conferred by transgene expression and accumulation of the product, whereas in other cases, resistance was reported due to dominant negative interference (Sinisterra et al., 1999)
use of virus movement proteins as sources of genes for resistance to virus infection was predicted following research in the late 1980s that identified viral genes that are responsible for local virus spread. The MP of TGMV shares 41% amino acid sequence identity with ACMV MP and it cannot complement the movement of ACMV. But it was found that in transgenic plants TGMV MP inhibited the spread of ACMV (Bejarano and Lichtenstein, 1994).

It is proposed that the dysfunctional MP prevents the accumulation of the movement protein of the challenge virus in such a way that the MP of the challenge virus is unable to complete its function (Beachy, 1999). The multifunctional Rep protein plays an integral role in regulation of viral gene expression as well as in initiation and termination of replication. These properties have made this protein a favorite target for PDR. Transformation of plants with plant viral replicase sequences, resulting in induction of resistance was first reported in 1990, employing the 54-kDa read-through region of the TMV replicase gene transformed into N. tabacum. The characteristic features of "replicase-mediated resistance" are the high degree of resistance shown both to viral and to RNA inocula, and the narrowness of the specificity (i.e., resistance is shown only to the virus from which the replicase gene was derived and to very closely related strains or mutants).

In most cases, there is a substantial inhibition of virus replication in initially inoculated cells and some limited cell-to-cell movement, but the infection does not spread from the inoculated leaf and no systemic disease develops (Nguyen, W.J.Ding, and B.Zaitlin, 1996). Similarly it was first shown (Hong and Stanley, 1996) that full length or truncated ACMV Rep inhibited the replication of ACMV in N. tabacum. But none of these transgenic plants was resistant to the quite distantly related viruses TGMV and BCTV (sharing 60% Rep amino acid sequence identity with ACMV), suggesting that the amino acid sequence was very important as far as resistance is concerned. This is not true for all cases. The molecular mechanism underlying replicase mediated resistance is not well understood as Tomato leaf curl New Delhi virus (ToLCNDV) resistant N. benthamiana expressing a truncated ToLCNDV Rep has shown a degree of cross-resistance to other begomoviruses such as ACMV (sharing 72% Rep amino acid sequence identity with ToLCNDV), Pepper huasteco
yellow vein virus (PHYVV, sharing 65% Rep identity) and Potato yellow mosaic virus (PYMV; sharing 64% Rep identity) (Chatterji, Beachy, and Fauquet, 2001). There does not therefore seem to be a simple sequence identity threshold beyond which Rep-based transgenes will not provide protection against viruses expressing distantly related Reps. Two possible risks associated with viral coat-protein expressing plants are heterologous encapsidation and recombination. Recombination between mutant viral genome and coat protein transgene can result in a permanent change in genetic constitution of the virus progeny resulting in infectious recombinant virus.

1.1.4 RNA-mediated pathogen derived resistance

Virus resistance can be engineered in plants by expressing transgenes that code for RNA species that are not translated into proteins. Four RNA-based strategies have been examined that result in the expression of a sense RNA, an antisense RNA, a defective interfering (DI) RNA, or a satellite RNA. Plants expressing transgenes that encode an antisense RNA (Day et al., 1991; Bejarano and Lichtenstein, 1994; Bendahmane and Gronenborn, 1997; Mubin et al., 2007), DI RNA (Rubio et al., 1999), a satellite RNA (Duan et al., 1997b), hairpin double stranded RNA (Asad et al., 2003; Pooggin et al., 2003) have been constructed and a minimal level of protection from virus-induced disease symptoms has been observed. RNA based protection strategy minimizes the risks associated with antiviral strategies. It has been proposed that protein product or the genetic material of the transgene being incorporated can be temporarily used by another viral pathogen to alter its host range or pathogenicity (Tepfer, 2002). Transcomplementation (Latham and Wilson, 2008; Froissart, Michalakis, and Blanc, 2002) transencapsidation (Pethybridge and Turechek, 2003), and recombination (Greene and Allison., 1994) have all been cited as potential problems.

The untranslatable sense RNA strategy negates concerns about transencapsidation and transcomplementation as no protein product is produced. RNA recombination between a viral genome and an untranslatable transgene transcript also should not pose a problem, as it should lead to a lethal recombination event. RNA mediated pathogen-derived resistance thus maximizes the effectiveness of the
technology. RNA-based resistance share many features in common: No transgene protein is required. Usually plant contains multiple copies of the transgene. It is associated with a high transcription rate but low steady state levels of transgene mRNA.

Plants are either resistant to virus infection (no detectable virus replication, spread or symptoms) or initially show virus infection and symptoms, but subsequently produce new growth that is symptom less and resistant to virus infection. Usually associated with methylation of transgene coding regions. Plants have resistance only to closely related virus strains (Waterhouse, Graham, and Wang, 1998). One of the underlying mechanisms for PDR is RNA silencing of the viral transgene (Lindbo et al., 1993).

1.14.5 Non-pathogen derived resistance

Genetically engineered resistance can be produced without involving pathogen genes. This includes incorporation of plant disease resistance genes from various hosts, ribosome inactivating proteins, plant proteinase inhibitors, antiviral antibodies expressed in plants, engineering for secondary metabolites conferring resistance, and systemic acquired resistance.

1.15 Other approaches for disease resistance

Incorporation of plant resistance genes is more durable than any other system. R genes encode proteins specific to the avirulence (avr) genes found in a pathogen, and initiate a cascade of signaling reactions after sensing the avr gene products that results in local cell death known as the hypersensitive response (HR). Actually, this strategy is based on the classical gene-for-gene hypothesis (Kaloshian, 2004). A number of structural features are conserved across several R gene products. These include a leucine-rich repeat (LRR), nucleotide-binding site (NBS), serine-threonine kinase, leucine zipper and toll-interleukin region (TIR) (Peart et al., 2005). The N gene of N. tabacum is well-characterized and the earliest known R gene which confers resistance against TMV. This gene was expressed in a heterologous host and a HR was produced when inoculated with TMV (Whitham et al., 1994).
**Introduction and Review of Literature**

*Turnip crinkle virus* (TCV) resistance in *A. thaliana* is mediated by an altogether different mechanism. The RTM gene, present in ecotype Columbia-O, brings about a HR independent resistance against TCV by affecting its long distance movement and is present as two alleles, RTM1 and RTM2 (Oingzhong et al., 1997). The RTM protein is believed to interfere directly with an essential component of the long-distance movement of the virus. Thus, model plants such as *Arabidopsis* can help us in looking for related R genes in crop plants (Whitham et al., 1994).

Several plants have been found to contain antiviral proteins known as ribosome inactivating proteins (RIPs). RIPs inhibit the translocation step of translation by catalytically removing a specific adenine base from 28S ribosomal RNA. They are synthesized either as pre- or pre-pro-proteins. The antiviral activity of several types of RIPs has been well-documented (Lam et al., 1996; Lodge, Kaniewski, and Tumer, 1993). When purified RIPs are mixed with viruses and applied on plants, virus multiplication and symptom development are dramatically suppressed. The development of systemic resistance was reported following studies on induction of a 34 kDa basic protein from the RIP (CA-SRI) treated *Cyamopsis tetragonoloba* plants.

In another experiment, the toxin gene, dianthin was placed downstream of a transactivatable geminivirus promoter from ACMV. When transgenic *N. benthamiana* plants were inoculated with ACMV, dianthin was synthesized only in the virus-infected tissues where it killed the cell (Hong and Stanley, 1996).

Another approach to control plant viruses is to express specific anti-viral antibodies, commonly known as plantbodies in plants. Single chain antibodies against the CP of *beet necrotic yellow vein virus* have been expressed in *N. benthamiana* and a significant delay in symptom development was reported, following mechanical inoculation (Joshi and Joshi, 1991).

Following viral infection, plants develop an active resistance, which is at first localized only at the site of infection, but spreads systemically in due course. This resistance, called systemic acquired resistance (SAR), is characterized by the coordinate activation of several genes in uninfected, distal parts of the inoculated plants. Tobacco plants were transformed with two bacterial genes coding for enzymes...
that convert chorismate into salicylic acid (SA) by a two-step process. When the two enzymes were targeted to the chloroplast, the transgenic plants showed a 500- to 1000 fold increased accumulation of SA and SA-glucoside, compared to control plants. The level of PR-proteins was enhanced and these plants showed resistance to viral and fungal infection, in a mode similar to SAR in non-transgenic plants (Ryals et al., 1996).

Metabolic pathways which are important in viral pathogenesis are key targets for intervention against viral infection. Another novel approach of interference with viral pathogenesis is to inhibit tetrapyrrole biosynthesis by expressing antisense RNA of uroporphyrinogen decarboxylase or coporphyrinogen oxidase in *N. tabacum*. The plants were characterized by accumulation of photosensitizing tetrapyrole intermediates, accumulation of highly fluorescent coumarin scopolin, PR proteins and reduced levels of infecting viral RNA.

### 1.16 RNA silencing

The central dogma of modern molecular biology, ‘DNA makes RNA makes protein’, predicts a role for RNA as a carrier of information, but it does not show RNA as a regulatory molecule. Although regulatory RNA had been observed in prokaryotes and eukaryotes, it has only recently emerged, with the discovery of RNA silencing, that it is a universal component of gene expression and control of gene expression. Still we don’t know much about RNA silencing. We have little knowledge about the mechanism of RNA silencing especially in plants due to non availability of mutants. But as research is going on all over the world we are now in a position to describe some of what RNA silencing can do, a little about how it does it, and we use it extensively as a research tool for gene knockdown through RNAi.

Andrew Z. Fire and Craig C. Mello were awarded noble prize in 2006 on the discovery of RNA silencing. RNA silencing, 2002’s “Technology of the year” by Science (Couzin, 2002) and Fortune Magazine's “Billion Dollar Breakthrough” (Stipp, 2003) in 2003, RNAi is a mechanism of suppressing gene expression by degrading specific messenger RNAs (mRNA). Introduction of a piece of dsRNA into the cytosol initiates the phenomenon of RNAi (Fig 1.8) in turn activating a pathway
Introduction and Review of Literature

culminating in the degradation of the targeted gene transcript (Agrawal et al., 2003; Zamore et al., 2000). In addition to RNA degradation upon activation of the RNAi pathway, there are also cases where the promoter region of the gene is silenced through methylation (Sijen et al., 2001; Wassenegger et al., 1994).

Although RNA silencing has only emerged as a topic of general interest in the past six years, the first RNA silencing paper was published as long ago as 1928. In that paper Wingard described tobacco plants in which only the initially infected leaves were necrotic and diseased owing to Tobacco ring spot virus (TRSV). The upper leaves had somehow become immune to the virus and consequently were asymptomatic and resistant to secondary infection. At the time this ‘recovery’ was a mystery; there was no obvious way to explain the specificity of the resistance to secondary infection (Wingar, 1928).

The details of the TRSV example remain to be worked out but we now know that recovery from virus disease involves RNA silencing that is targeted specifically at the viral RNA (Covey and Al-Kaff, 2000). There was no information about mechanisms in 1928; it was not even known that the viral genome is RNA. But Wingard’s paper is an appropriate starting point for the current interest in RNA silencing because it illustrates a defense role against viruses for RNA silencing which may have been one of its original functions in primitive eukaryotes.

In plants this process has diversified into mechanisms that, in addition to defending the plant against viruses, protect the genome from transposons and regulate gene expression (Baulcombe, 2004). RNA silencing involves suppression of gene expression by sequence-specific interaction with RNA either at the transcriptional or the posttranscriptional level in diverse eukaryotes.

The RNA silencing phenomenon was termed post-transcriptional gene silencing (PTGS) in plants (Van Rij, 2006), quelling in fungi (Escobar et al., 2001) and RNAi in animals (Fire et al., 1998). RNA silencing is a complex phenomenon which results in the sequence specific inhibition of gene expression. This inhibition could be at the transcription level, might be target stability of RNA or at the translational level (Baulcombe, 2004). These processes share three biochemical
features. RNA formation, second dicing of dsRNA to small 20–26-nt dsRNAs with two bases overhangs and third inhibitory action of a selected sRNA strand within effector complexes called the RNA-induced silencing complex (RISC) acting on partially or fully complementary RNA or DNA (Itaya et al., 2007). One of the two small (sRNA) strands (antisense) joins (RISCs), which have a number of multifunctional proteins called Ago proteins. Agos have a sRNA binding PAZ domain and also contain a PIWI domain that provides endonucleolytic (‘slicer’) activity to those RISCs programmed to cleave target RNAs. These Agos are very critical for the RNA silencing (Parker, Roe, and Barford, 2004).

1.17 RNA silencing through the exogenous introduction of transgene

1.17.1 Post-transcriptional gene silencing (PTGS)

PTGS was discovered in transgenic Petunia when in order to get more color researchers tried to over express a gene oncoding chalcon synthetase, which result in as loss of expression of both the transgenes and homologous endogenous genes. But if we express the transgene in both sense and anti-sense orientation together the efficiency of PTGS can be enhanced (Waterhouse, Graham, and Wang, 1998) or by direct production of long dsRNA from inverted-repeat (IR) transgenes having an intron. Now-a-days, most of the labs in the world are using the IR-PTGS, as an experimental RNAi in plants. Two distinct sRNA classes termed short interfering siRNAs are produced in this phenomenon. 21 nt siRNAs are believed to guide mRNA cleavage, whereas 24 nt siRNAs are believed to exclusively mediate chromatin modifications (Hamilton et al., 2002).

Both siRNA classes accumulate as populations along the entire sequence of IR transcripts (Llave et al., 2002). Here a question arise that whether there are structural or functional reasons that make 24nt sRNA more suitable for chromatin modification than 21nt sRNA and less well suited to mRNA cleavage. This is not the case because there are many reports that these sRNAs can perform both function. For example 21nt sRNA are fully competent to guide the methylation of a transgene. But during the course of evolution different sRNAs have been assigned different functions. RNA silencing especially IR-PTGS is very complex phenomenon. Despite of the efforts of
many laboratories around the world we are still unable to recover any mutant defective in the IR-PTGS pathway. So our understanding of this process is not very detailed (Béclin et al., 2002). How are the players of RNA silencing such as Dicers and RISCs activated in IR-PTGS?

One possibility could be the high level production of dsRNA which promotes the activities of different Dicers and RISCs, which would normally act in distinct pathways, act to mediate silencing redundantly. There are many reports in which people have analysed combinatorial Dicer knockouts in Arabidopsis (Xie et al., 2005). However, Dicer-like 4 (DCL4) seems a preferred enzyme for IR-PTGS: it was specifically required for 21 nt siRNA accumulation and silencing from a moderately expressed, phloem-specific IR transgene (Dunoyer and Voinnet, 2005). The possibility of the involvement of DCL2 cannot be ruled out as it processes some endogenous DCL4 substrates into 22 nt long siRNAs in the absence of DCL4 (Xie et al., 2005) although it remains unclear whether the 22 nt molecules can functionally substitute for the 21-nt siRNA products of DCL4 (Deleris et al., 2006). Another dicer enzyme, DCL3, might be involved in the processing of 24nt short RNAs which can direct DNA or histone modifications. The function of DCL3 is not essential for cleavage of dsRNA. both these types of siRNAs undergo HEN1 mediated methylation at their 3’ termini. This methylation protects these sRNAs from oligo-uridylation, a modification that promotes instability of these sRNAs (Yang et al., 2006).

1.17.2 Small RNA molecules

A microRNA (miRNA) is a 21–24 nucleotide small RNA that is the final product of a non-coding RNA gene. The miRNA genes resemble protein coding genes in that they may contain introns and that they are transcribed by RNA polymerase II. Like other pol II transcripts, the transcripts from miRNA genes are capped, spliced and polyadenylated (Kim, 2005). The mature miRNA is located in a hairpin structure within the primary transcript (pri-miRNA) and is processed from the pri-miRNA through at least two RNase III-mediated steps (Bartel, 2004). The miRNA is loaded into RISC, where it guides the cleavage or translational repression of its target mRNAs by base-pairing with the targets. Plants are rich in another type of small RNAs (21-24) (known as siRNAs) that is similar in structure, biogenesis and function.
to miRNAs. The size of sRNA, whether 21 or 24nt, depends upon the dicer enzyme which process these sRNAs. siRNAs originate from transcripts from transgenes (Hamilton and Baulcombe, 1999) endogenous repeat sequences or transposons (Hamilton et al., 2002).

One key distinction between miRNAs and siRNAs from transgenes and repeat sequences/transposons is that miRNAs target genes other than the ones that give rise to the miRNAs while siRNAs target the very sequences that generate them (Bartel, 2004). A new class of siRNAs that, like miRNAs, targets mRNAs from other loci was also discovered in Arabidopsis and named trans-acting siRNAs (ta-siRNAs). The ta-siRNAs originate from loci that give rise to non-coding transcripts that are themselves targets of miRNAs. The miRNA-mediated cleavage of the transcripts recruits an RNA-dependent RNA polymerase (RdRP) to use the cleaved transcripts as templates to generate long dsRNAs, which then serve as the source of multiple ta-siRNAs. The miRNA-mediated cleavage of the precursor RNA is crucial for the biogenesis of the ta-siRNAs and sets the register for the cleavage events that generate the ta-siRNAs. Since ta-siRNAs target mRNAs from other genes, the cis/trans relationship between small RNAs and their targets is no longer a distinction between miRNAs and siRNAs. Now it seems that the only feature that distinguishes miRNAs and siRNAs is the nature of the precursor transcripts. While a miRNA comes from a hairpin pre-miRNA, siRNAs come from a perfect, long dsRNA generated by an RdRP or through the transcription of a hairpin transgene. Usually only one miRNA is generated from the pre-miRNA but several or many siRNAs are generated from long dsRNAs.

However, there is a case in which more than one small RNA molecule comes from a single pre-miRNA (Kurihara and Watanabe, 2004). A synthetic mi/siRNA has been made using the primary miR171 transcript but in which the miRNA sequence has been replaced by a sequence that targets GFP mRNA. Another difference between miRNA and siRNA is that siRNAs are double stranded but miRNAs are single stranded. Drosha–Pasha and Dicer-1-Loquacious complexes are involved in the processing of miRNAs while siRNAs are processed by DCL2, DCL3, RDR6, SGS3 and SDE3 (Fusaro et al., 2006).
Recently, a third class of sRNAs been shown to be produced by *Drosophila* which is termed endogenous small interfering RNAs (Hammond et al., 2000). Production of these RNAs requires Dicer-2, but a subset depends preferentially on Loquacious rather than the canonical Dicer-2 partner, R2D2. These endogenous small interfering RNAs predominantly join Argonaute-2 and have the capacity, as a class, to target both protein-coding genes and mobile elements. These observations blur the distinctions among the short RNA classes (Kraynack and Baker, 2006).

The biogenesis of miRNAs in *Arabidopsis* is similar to that in animals in that miRNAs are processed from primary precursors by at least two steps mediated by RNase III-like enzymes and that the miRNAs are incorporated into a protein complex named RISC (Carrington and Ambros, 2003). However, the biogenesis of plant miRNAs consists of an additional step, i.e. the miRNAs are methylated on the ribose of the last nucleotide by the miRNA methyltransferase HEN1 (Zhang et al., 2007).

The high degree of sequence complementarity between plants miRNAs and their target mRNAs has facilitated the bioinformatics prediction of miRNA targets, many of which have been subsequently validated. Plant miRNAs have been predicted or confirmed to regulate a variety of processes, such as development, metabolism, and stress responses. A large category of miRNA targets consists of genes encoding transcription factors that play important roles in patterning the plant form (Chen, 2005).
Figure 1.8 A model depicting geminivirus-induced RNA silencing in plants. The geminiviral replication cycle is shown in the nucleus of the cell. RNAi is triggered by dsRNAs generated from overlapping, abundant and folded forms of mRNAs. An enzyme called DICER cleaves long dsRNA into siRNAs. An RNA-induced silencing complex (RISC) then distinguishes between the different strands of the siRNA. The sense-strand (depicted in blue) is degraded. The anti-sense strand (depicted in red) is used to target genes for silencing. The anti-sense strand bound to an RdRP enzyme, can pair with a complementary mRNA and act as a start point for the synthesis of a new long dsRNA. DICER is then required to generate transitive siRNA (depicted in green), which are specific to different sequences on the same mRNA. All these processes lead to mRNA degradation. This diagram is based on figure 1 in Vanitharani, Chellappan, and Fauquet (2005).
Chapter 2
2.1 DNA Manipulation

2.1.1 Sample collection

Symptomatic plants were collected from farmer’s fields after the plants were photographed with a high resolution digital camera. Young leaves of symptomatic plants were collected, kept in plastic bags labelled with permanent marker and transported on ice before being stored at -80°C until they were utilized for DNA extraction.

2.1.2 DNA extraction from plant samples

DNA was extracted from leaf samples by the CTAB method described by Doyle and Doyle (1990). 100 to 200mg of leaf tissue was ground in liquid nitrogen in a pestle and mortar. In a microcentrifuge tube the powdered tissue was mixed with 700µL of pre-warmed CTAB buffer (100mM Tris-HCl [pH 8.0], 20mM EDTA, 1.4M NaCl, 2% [w/v] Cetyl Triethyl Ammonium Bromide (CTAB) and 0.02% (v/v) β-mercaptoethanol) and incubated at 65°C for 30 minutes.

After lowering the temperature of the samples to room temperature, an equal volume of chloroform/isoamyl alcohol (24:1) was added, mixed well and centrifuged at 9000 rpm for 10 min at room temperature in a microfuge (Eppendorf model 5414D). The upper aqueous, DNA-containing phase was taken into a new microcentrifuge tube and mixed with 0.6 volume isopropanol to precipitate the DNA. DNA was pelleted by centrifugation at 13,200 rpm for 10 minutes and, after discarding the supernatant, the pellet was washed with 70% (v/v) ethanol and air dried. Finally the pellet was dissolved in sterile distilled water (SDW).

2.1.3 Quantification of DNA

DNA concentrations were determined using a spectrophotometer (SmartSpec Plus, BIORAD) by taking the absorbance at a wavelength of 260nm with conversion factor of 1 OD = 50 µg/ml.
2.2 Amplification of DNA

2.2.1 PCR amplification of DNA

For amplification of DNA by PCR a reaction mixture of 50µL containing 10pg-1µg template DNA, 5µL 10X Taq polymerase buffer (Fermentas), 5µL 2mM dNTP mix., 1.5mM MgCl₂, 0.5µM each of primers and 1.25u Taq DNA polymerase (Fermentas) was prepared in 0.25mL or 0.5mL PCR tubes. The reaction mixture was incubated in thermal cycler (Eppendorf Model AG22331, Hamburg, Germany or Creacon technologies Model 0005.400, Netherlands).

The thermal cycler was programmed for a preheat treatment of 94°C for 5 minutes followed by 35 cycles of 94°C for 1min, 48°C to 52°C for 1min and 72°C for varying times (dependent upon the length of fragment to be amplified; typically 1 min per 1000 nucleotides to be amplified), followed by a final incubation of 10 min at 72°C and routinely the machine was set to hold at 4°C until the samples were removed. Universal as well as specific primers were used for DNA amplification from all samples. For diagnostic PCR the volume of the reaction mixture was reduced to 25µL per tube by reducing the ingredients accordingly.

2.2.2 Rolling-circle amplification (RCA)

For amplification of circular DNA molecules by RCA a reaction mixture of 20µl containing 100 to 200ng genomic DNA of infected plant samples, 50 µM random hexamer primers, 2µl 10X Φ29 DNA polymerase reaction buffer (330mM Tris-acetate [pH 7.9] ), 100mM magnesium acetate, 660mM potassium acetate, 1% (v/v) Tween 20, 10mM DTT) was prepared and incubated at 94°C for 3 minutes to denature double stranded DNA.

The mixture was cooled to room temperature and mixed with 1mM dNTPs, 5-7 units of Φ 29 DNA polymerase and 0.02 unit of pyrophosphatase (to eliminate inhibitory accumulation of pyrophosphate) and incubated at 30°C for 18 to 20 h. The following day Φ 29 DNA polymerase was inactivated at 65°C for 10 minutes.
Materials and Methods

2.3 Cloning of amplified DNA

2.3.1 Cloning of PCR product

PCR amplified DNA was cloned using the InstaClone PCR Cloning Kit (Fermentas) according to the instructions given by the manufacturer. In brief, a reaction mixture of 30µL containing 18 to 540ng PCR product (depends upon the length of DNA fragment), 3µL vector (pTZ57R/T), 6µL 5X ligation buffer and 5u T4 DNA Ligase, was prepared in a 1.5mL microcentrifuge tube and incubated at 16°C overnight. The following day the ligation mixture was transformed to competent cells of E.coli (DH5α) by the heat-shock method.

Solid LB media plate with 100µg/mL ampicillin, spread with 20µL (50mg/mL) X-Gal and 40µL (20mg/mL) IPTG were used for plating-out the transformed cells and incubated at 37°C for 16 hours. On appearance of colonies white colonies were picked by sterile tooth picks, inoculated in 5mL LB (1% [w/v] tryptone, 0.5% yeast extract and 1% [w/v] NaCl) liquid media in autoclaved test tube and grown at 37°C in a shaker overnight. The following day plasmids were isolated from E. coli cultures by the miniprep method and screened by restriction analysis.

2.3.2 Cloning of RCA product

RCA product was digested with unique cutter enzymes into monomers. Cloning vector (usually pBluescript II KS(SK [-]) was also restricted with same enzyme. Restricted RCA product and vector were treated with phenol-chloroform or pass through the column, using the kit by Wizard SV Gel and PCR Clean-Up System (Promega) by the method described by the manufacturer to remove salts.

Vector and insert were ligated in a reaction mixture of 20 µL containing vector and insert in 1:3 ratios, 4µL 5X ligation buffer and 1µL T4 DNA ligase. Ligation mixture was kept at 16°C overnight and next day transformed into competent E. coli cells.
2.4 **Transformation of competent cells**

2.4.1 **Transformation of heat-shock competent *E. coli* cells**

Transformation of competent *E. coli* cells was carried out by the methods described by Sambrook, Frisch and Maniatis (1989). The ligation mixture was added to thawed competent *E. coli* (200ul), mixed gently and incubated on ice for 30min. The cells were shocked at 42°C a dry bath/water bath. After 1-2 minutes cells were transferred to ice and incubated for two minutes. 1mL LB liquid medium was mixed in each tube and put on shaking at 37°C for 1h. Transformed cells were spread on solid LB media plates with appropriate antibiotics and kept at 37°C in an incubator overnight.

2.4.2 **Transformation of competent *Agrobacterium tumefaciens* cells**

2µL of the plasmid was mixed with electro-competent *A. tumefaciens* cells (LBA4404 or GV3101) on ice and transferred to a chilled electroporation cuvette. Electroporator (BTX Harvard Apparatus) was set at 1.44kV. The cuvette was inserted into the electric shock chamber and start button was pressed. After the shock, LB liquid medium was added to the cells and the tubes incubated at 28°C for 2h in a shaker. Electroporated cells were spread on LB medium with appropriate antibiotics kanamycin (30ug/ml); rifampicin (50ug/ml) and tetracycline 12ug/ml and Chlorophenicol ( 20ug/ml (for LBA4404 , for GV3101, for pGreen0029, pGSA1403) wrapped with aluminium foil and incubated at 28°C for 48h.

2.4.3 **Plasmid isolation**

Using a sterile tooth pick, a single bacterial colony from a plate was picked and inoculated into 5mL LB broth with appropriate antibiotic selection in a sterile culture tube. It was incubated overnight at 37°C with shaking at 250 rpm. When the culture was ready, it was decanted in 1.5mL microcentrifuge tubes and centrifuged at full speed for two minutes to harvest the cells.

The supernatant was discarded and the pellet was re-suspended in 100µL Resuspension solution (50 mM Tris-HCl [pH 8.0], 10 mM EDTA, 100 ug/mL RNase A) using a vortex. 200µL Lysis solution (0.2 M NaOH, 1% [w/v] SDS) was added and
Materials and Methods

mixed gently. 200µL Neutralization solution (3.0M Potassium Acetate, 5.0M glacial acetic acid, [pH 5.5]) was added, mixed thoroughly and centrifuged at 13,200 rpm for 10 minutes.

The supernatant was transferred to a new microcentrifuge tube. Two volume (approximately 1mL) chilled absolute ethanol was added to precipitate the DNA and centrifuged for 10 minutes to pellet the DNA. DNA pellet was washed with 70% (v/v) ethanol, air dried and dissolved in sterile distilled water (SDW).

For DNA sequencing, the plasmid was isolated by GeneJET Plasmid Miniprep Kit (Fermentas). The culture of *E. coli* was decanted into 1.5 mL microcentrifuge tube and centrifuged for 2 minutes. The pellet was re-suspended in 250µL resuspension Solution and cells were lysed with 250µL Lysis Solution. 350µL Neutralization Solution was added, mixed thoroughly and the tube was centrifuged at 13,200 for five minutes. A mini-column provided with the kit was inserted into the collection tube and the supernatant was transferred to the column. The column was centrifuged for one minute to bind the DNA to the matrix and the flow through in the collection tube was discarded. The matrix was washed twice, first with 700µL Column Wash Solution was added and incubated at room temperature for one minute and then centrifuged for one minute. The Wash solution was removed from the column and the column washed with 500µl Column Wash solution. Subsequently the column was centrifuged for one minute with an empty collection tube to remove residual ethanol. Finally the column was inserted into a fresh microcentrifuge tube. DNA in the column was dissolved in 50µL SDW, incubated at room temperature and recovered by centrifugation.

2.4.4 Digestion of plasmid DNA

Digestion of plasmids and PCR products was done using restriction endonucleases and their corresponding buffers in accordance with the supplier’s (Fermentas, Biolab or Promega) guidelines. A total reaction volume of 10µl was used when screening plasmid preparations for the expected insert, 20µl for digestions incubated overnight for cloning.
2.5 DNA Analysis

2.5.1 Agarose-gel electrophoresis

DNA was mixed with 5X loading dye and electrophoresed in 1% (w/v) agarose gels containing ethidium bromide (0.5μg/ mL). Gels were prepared in a minigel apparatus (12 x 9 cm) or midigel apparatus (18 x 15 cm), containing either 1X TBE (890mM Tris [pH 8.3], 890mM boric acid, 20mM EDTA,) or 1X TAE (40mM Tris-acetate [pH 8.4], 1mM EDTA) buffer. TBE gels were electrophoresed at approximately 50V and TAE gels at 110V. The DNA was viewed using a short wavelength ultraviolet (UV) transilluminator (Eagle Eye-Stratagene) and fragment length estimated by comparison with a co-electrophoresed 1 kbp DNA ladder (Fermentas).

2.5.2 Southern blot analysis

10µg genomic DNA per well was loaded on a 1% (w/v) agarose gel and run at 40V (Biorad PowerPac™) in TBE buffer for 4 to 5 h. Gel was stained with 0.5μg/mL ethidium bromide and DNA image was obtained under UV light in gel documentation apparatus (Eagle Eye-Stratagene). After electrophoresis the gel was treated with depurination solution (0.25M HCl) for 15 minutes, denaturation solution (1.5M NaCl and 0.5M NaOH) for 30 minutes and neutralization solution (1M Tris [pH 7.4], 1.5M NaCl) for 30 minutes. The gel was rinsed briefly with distilled water between treatments and shaken moderately on platform shaker during each treatment. DNA in the gel was transferred to a nylon membrane (Hybond-Amersham) in 10X SSC and sometimes in 5X SSC (1.5M NaCl and 150mM sodium citrate) by capillary action. The apparatus used to blot the gel is shown in (Fig. 2.1). The DNA on the nylon membrane was crosslinked by UV irradiation (CL-1000 , UVP) at 120mJ/cm2 energy. The membrane was then rinsed in a solution containing 0.1X SSC, 0.5% (w/v) SDS at 65°C for 45 minutes to remove residual agarose.

Before hybridization the membrane was treated with 0.2 ml/cm² pre-hybridization solution (6X SSC, 5X Denhardt’s solution [0.1% (w/v) each of bovine serum albumin, 0.5% (w/v) Ficoll (Mol. Wt. ~400,000) and PVP (Mol. Wt. ~40,000)], 50% (v/v) SDS) and 5mg/mL sheared and denatured salmon sperm DNA at 42°C for 2–4 hours in a hybridizer (Hybaid, Midi Dual 14), to block non-specific binding sites.
Materials and Methods

DNA probes were prepared using a Biotin DecaLabel DNA Labelling kit (Fermentas) according to the manufacturer’s instructions. Briefly, in a 1.5mL microcentrifuge tube a 44 µL reaction mixture was prepared by adding 50-200ng DNA template (usually purified PCR product), 10µL decanucleotide in 5X reaction buffer and nuclease free water. The reaction mixture was vortexed briefly, centrifuged briefly in a microfuge to collect the contents at the bottom of the tube and incubated in a boiling water bath for 5-10min. After incubation the tube was cooled on ice, briefly microfuged and the contents of the tube mixed with 5µL biotin labelling mixture and 1µL Klenow fragment exo- (5units) and incubated at 37°C for 1 hour to 20 hour. Reaction was stopped by adding 1µL 0.5M EDTA [pH 8.0]. To prepare hybridization solution, the biotin labelled probe was denatured at 100°C for 5 minutes, chilled on ice and mixed with pre-hybridization solution (25-100ng/mL).

After 2-4h treatment, the pre-hybridization solution was discarded and the hybridization solution was added to the membrane (60µL/cm2) and incubated overnight in a hybridizer at 42°C. The following day the membrane was washed twice with 2X SSC/0.1% (w/v) SDS at room temperature for 10 minutes. The membrane was washed with 0.1X SSC/0.1% (w/v) SDS twice at 65°C for 20 minutes. To detect the biotin-labelled DNA the membrane was washed in 30mL Blocking/Washing Buffer (provided by the manufacturer) at room temperature. After 5 minutes the membrane was treated with 30mL Blocking Solution for 30 minutes to block non-specific binding sites on the membrane. Streptavidin-AP conjugate was diluted in 20mL Blocking Solution and the membrane was incubated in it for 30 minutes. The membrane was washed twice in 60mL Blocking/Washing buffer for 15 minutes and incubated with 20mL Detection Buffer for 10 minutes. Finally the membrane was treated with 10mL freshly prepared Substrate Solution at room temperature in the dark until blue-purple precipitate became visible. To stop the reaction, the substrate solution was discarded and the membrane was rinsed with water. The blot was immediately photographed and the membrane was then air dried and stored.
Materials and Methods

Figure 2.1 Capillary blotting apparatus for the transfer of nucleic acids from agarose gels to Hybond membrane.

2.6 Purification of DNA

2.6.1 Gel extraction and PCR product purification

DNA was run on 1% (w/v) agarose gels and the desired fragments were cut out from the gel using a scalpel under UV light. DNA from the gel was isolated using a Wizard SV Gel and PCR Clean-Up System (Promega) by the method described by the manufacturer. The excised gel slice was weighed and placed in a 1.5mL microcentrifuge tube to which was added 10µL Membrane Binding Solution per 10mg of gel slice. The tube was vortexed and incubated at 55-65°C until the gel slice was completely dissolved. An equal volume of Membrane Binding Solution was mixed with the dissolved gel mixture and then transferred to the Minicolumn assembly, incubated at room temperature for 1 minute and centrifuged at 16,000×g for 1 minute. The flow-through was discarded and the Minicolumn was reinserted into the collection tube. 700µL of Membrane Wash Solution was added and the tube with column centrifuged at 16,000×g for 1 minute. Again the flow-through was discarded and 500µL Membrane Wash Solution was added in the column. After 5 minutes of centrifugation at 16,000×g the collection tube was emptied and the Minicolumn, with
Materials and Methods

empty collection tube, was centrifuged for 1 minute with the lid open to allow evaporation of any residual ethanol. Finally the Minicolumn was transferred to a clean microcentrifuge tube, 50µL nuclease-free water was added to the Minicolumn, incubated at room temperature for 1 minute and centrifuged at 16,000×g for 1 minute. The Minicolumn was discarded and purified DNA was stored at -20°C.

2.6.2 Phenol-chloroform treatment of DNA

To remove proteins from, phenol:chloroform (1:1) extraction method was used. Equal volume of phenol:chloroform was mixed with the DNA solution and vortexed until the mixture turned milky. It was centrifuged at 13,200 rpm for 10 minutes and upper aqueous phase was collected in such a way that the layer between upper aqueous phase and phenol:chloroform was not disturbed. 1/10 volume 3M sodium acetate [pH 5.4] and 2.5 volume chilled absolute ethanol was mixed into supernatant and placed at -20°C in a freezer for one hour. To pellet the DNA, it was centrifuged at maximum speed. DNA pallet was washed with 70% (v/v) ethanol, air dried and dissolved in appropriated amount of SDW.

2.7 Microbiological Techniques

2.7.1 Preparation of heat shock competent Escherichia coli cells

A single colony from a freshly grown plate of E. coli was inoculated into 20mL LB medium in a 50mL flask and incubated at 37°C overnight with vigorous shaking. The following day 2mL of the overnight culture was taken and diluted to 250mL in 1L flask and shaken vigorously at 37°C until an OD₆₀₀ of 0.5-1 was achieved.

The culture was chilled on ice for 30 minutes, transferred aseptically to sterile disposable 50mL propylene tubes and centrifuge at 4000 rpm at 4°C for 5 minutes to pellet the cells. The pellet was resuspended in 20mL of 0.1M MgCl₂ and centrifuged again. Pallet was re-suspended in 20mL of 0.1 M CaCl₂, incubated on ice for 30 minutes and centrifuged at 400 rpm. Finally the pellet was re-suspended in appropriate amount of 0.1M CaCl₂ and filter-sterile cold 30% (v/v) glycerol. The cells were stored in aliquots of 200µl at –80°C.
2.7.2 Preparation of electro competent *Agrobacterium tumefaciens* cells

Before preparing the electro-competent cells, the *Agrobacterium tumefaciens* strain GV3101 was transformed with pSoup plasmid and grown in LB media containing tetracycline (50 µg/ml) as the pG0029 can’t replicate in *Agrobacterium* without the pSoup being co-resident (Hellens, Mullineaux, and Klee, 2000). A single colony from a freshly grown plate of *A. tumefaciens* was transferred using a sterile tooth pick and inoculated into 20mL LB liquid medium with 25µg/mL rifampicin in a 50mL autoclaved flask and incubated with vigorous shaking (140 rotations per minute) at 28°C for 48 hours.

5mL of the culture was re-inoculated into a 1L flask containing 250mL-350mL of the LB medium with 25µg/mL rifampicin and put on shaking at 28°C until the OD$_{600}$ of the cells was 0.5-1. When the desired OD$_{600}$ of the cells was achieved, the flask was cooled on ice for half an hour before starting the next step. Set the centrifuge temperature of about 4°C before starting the cells making. The cells were transferred aseptically to ice cold 50mL propylene tubes, centrifuged at 4000 rpm for 10minutes at 4°C. The pellet was re-suspended in 50mL of cold SDW and centrifuged again under the same conditions. Cells were again re-suspended in cold double distilled deionized water and the wash was repeated.

Now the cells were re-suspended in 10mL cold SDW containing filter sterilized 10% (v/v) glycerol and centrifuged at 4000 rpm. This step was repeated. Finally the cells were re-suspended in 2ml of filter sterilized cold 10% (v/v) glycerol, aliquoted in 1.5mL microcentrifuge tubes and stored at –80°C.

2.7.3 *Agrobacterium*-mediated inoculation

Clones in the binary vector pGreen0029 were electroporated to *Agrobacterium* strain GV3101. For agro-inoculation, glycerol stocks of *Agrobacterium* strains GV3101 with required clones were streaked on solid LB plates containing 12.5µg/mL rifampicin and 50µg/mL kanamycin and incubated at 28°C for 48h. A single colony of bacterial cells was transferred with a sterile wire loop and inoculated in 50mL liquid media containing antibiotics and put on shaking (160 rotations per minute) at 28°C until the O.D$_{600}$ of the culture was 1.
The cells were harvested by centrifugation at 4000 rpm for 15 minutes. The supernatant was discarded and the pellet was resuspended in 10mM magnesium chloride containing 100µM acetosyringone. For agro-inoculation to *N. benthamiana* and *N. tabacum*, plants at 4 to 5 leaves stage were not watered for 24 hours before inoculation. Tomato, beans, squash, pumpkin, cotton, *N. tabacum* and *Nictotiana benthamiana* seedlings were inoculated by slightly puncturing the leaf with a clean pin and infiltrating the bacterial suspension under pressure into the leaf stab using a sterile disposable 5ml syringe without a needle.

### 2.7.4 Plant growth conditions

All plants were grown in controlled conditions in growth rooms at 25°C with 16h dark period/8h light period and 65% humidity in small 5 inch diameter plastic pots containing clay, silt, sand and compost in equal proportions. Cotton plants were grown at 28°C. All plants were watered daily and with Hoagland solution by Merck™ (60% Foliar Lettice, 10% Linear Sulphate, 0.37% w/w Iron Chelate, 9.57% w/w Nitrogen, 0.01% Catalytic Enzyme Glyosides, Artificial colouring.) once a week.

### 2.7.5 Sequencing and sequence analysis

Plasmids were purified using a GeneJET Plasmid Miniprep Kit (Fermentas) and their sequences were determined commercially by Macrogen (South Korea). Sequences at the ends of clones were determined using the M13Forward (-20) and M13Reverse (-20) primers. These sequences were then extended by designing specific primers; a process known as primer walking.

Sequence data were assembled and analysed with the aid of Lasergene package of sequence analysis software (DNAStar Inc., Madison, WI, USA). A sequence similarity search (Blast) was performed by comparing the sequence to other begomovirus sequences in the database (http://www.ncbi.nlm.nih.gov/BLAST/) and open reading frames were located using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Final sequences were submitted to the EMBL sequence database. Multiple sequence alignments were performed using Clustal X (Thompson et al., 1997) and MegAlign program of the Lasergene package.
Materials and Methods

Phylogenetic trees were constructed using the Neighbour Joining algorithm of Clustal X and displayed, manipulated and printed using Treeview (Page, 1996).

2.7.6 Photography and image processing/manipulation

A digital camera (Sony Model DSC- W50) was used to photograph the infected plants in the field as well as in the glasshouse. The same camera was used to photograph the GFP fluorescence under ultraviolet light using a hand-held UV lamp (UV semiconductor inspection lamp products, Upland , CA; Black Ray model B 100AP). The photographs were manipulated with Adobe Photoshop CS. The figures were produced using CorelDRAW 12.
## Table 2.1 Sequences of primers used during the study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA101</td>
<td>CTGCAGATAATGCTAGTTACCAG</td>
</tr>
<tr>
<td>DNA102</td>
<td>CTGCAGATCCTCCACGTGATAG</td>
</tr>
<tr>
<td>DNA1mutF</td>
<td>AAAATATCGATGGTACCCTTGAGGAAGG</td>
</tr>
<tr>
<td>DNA1mutR</td>
<td>TATTAATCGATTATATCCATATATCCGCC</td>
</tr>
<tr>
<td>GFP F</td>
<td>ATCGATGGTGATGGTATGGAC</td>
</tr>
<tr>
<td>GFP R</td>
<td>ATCGATGTCATGCGGGTTCATATG</td>
</tr>
<tr>
<td>One F</td>
<td>GAAACTAGGTTTCTCRGGRC</td>
</tr>
<tr>
<td>One R</td>
<td>ATACCTTCCATTACGCGAAGG</td>
</tr>
<tr>
<td>DNA1’F</td>
<td>CGTAAGCTTAATCTGCCAGAGGTGAGGGGGG</td>
</tr>
<tr>
<td>DNA1’R</td>
<td>CGAAAGCTTTGACCGTGAGATGTGTGTT</td>
</tr>
<tr>
<td>DNA1”F</td>
<td>GACGTTCAGGATTTCCAG</td>
</tr>
<tr>
<td>DNA1”R</td>
<td>GATTCCCTTAGCTAGCAAGTC</td>
</tr>
<tr>
<td>Gh PDS F</td>
<td>GACACCGCGTGGATGGAACCCGTCCC</td>
</tr>
<tr>
<td>Gh PDS R</td>
<td>ATGACCGCGTICTCCAGGTTCTCAG</td>
</tr>
<tr>
<td>rs GFP F</td>
<td>ATGACCGCGTAGTAAAGGGAGAAGAC</td>
</tr>
<tr>
<td>rs GFP R</td>
<td>CGTACCGCGTCAGGCTGGCAGTGGCGCTCT</td>
</tr>
<tr>
<td>PDS48F</td>
<td>TTTTGCGATGCCATGCTGCA</td>
</tr>
<tr>
<td>PDS250R</td>
<td>TCTCTCCAGTTCTCCAGGCCA</td>
</tr>
<tr>
<td>Chl182F</td>
<td>CTGGGCTGTCAGGAGCC</td>
</tr>
<tr>
<td>Chl427R</td>
<td>TTCCGAAATCGAATCTGAAGCGG</td>
</tr>
<tr>
<td>GFP56F</td>
<td>TCAAGGAGCCGAAGGAACATC</td>
</tr>
<tr>
<td>GFP785R</td>
<td>AAAGGGGAGATTGTGTTGGAC</td>
</tr>
<tr>
<td>DNA1 Rep F</td>
<td>TAAAAGCTTTATGCTACTATTCCGTC</td>
</tr>
<tr>
<td>DNA1 Rep R</td>
<td>TTAGAAATTCTTATTCCATATATCCGC</td>
</tr>
<tr>
<td>Iteron1F</td>
<td>AAGCTT CTCTGGGTGTCCCTGAGTC</td>
</tr>
<tr>
<td>Iteron1R</td>
<td>AAGCTT AAGGGCTGGTGAATACTACACC</td>
</tr>
<tr>
<td>Iteron2F</td>
<td>AAGCTTGTGCCCTTACGTGCTTCTC</td>
</tr>
<tr>
<td>Iteron2R</td>
<td>AAGCTTAAAGGGAGAGGAACCCGAG</td>
</tr>
<tr>
<td>Rub F</td>
<td>CCAATCGATCCATGGATGTC</td>
</tr>
<tr>
<td>Rub R</td>
<td>TTTATCGAGTAGTTAATACACC</td>
</tr>
<tr>
<td>Pk3repF</td>
<td>CCCGGGGCTGGTCCCTCCATTGACCAAAAGTC</td>
</tr>
<tr>
<td>Pk3repR</td>
<td>CCCGGGGCGTGATCCCATGGTGCTTG</td>
</tr>
<tr>
<td>Begomo. F</td>
<td>ACCCCTG GCCGTGTCCGTCGCCCTACGTGACA</td>
</tr>
<tr>
<td>Begomo. R</td>
<td>ACACGGT ATGGGCTGYCAGAATATGACAG</td>
</tr>
<tr>
<td>β01</td>
<td>GTTACCACACTAGCCTACGCAGGAGGCC</td>
</tr>
<tr>
<td>β02</td>
<td>GTTACCTACCCTCCAGGAGGTTACA</td>
</tr>
<tr>
<td>TLCVBF</td>
<td>GCTAAGCTTCTGCTCGAACATGGAGATGGAA</td>
</tr>
<tr>
<td>TLCVBR</td>
<td>CAGAAGCTTAGCCAGGTAGGAAATAG</td>
</tr>
</tbody>
</table>
Chapter 3
3.1 Introduction

Recently a group of begomoviruses have been identified that are monopartite but require a satellite molecule (collectively known as betasatellites) to induce typical disease symptoms in the plant species from which they were isolated (Briddon et al., 2001; Jose and Usha, 2003; Saunders, Bedford, and Stanley, 2002; Saunders et al., 2003; Zhou et al., 2003). Betasatellites (previously known as DNA β; Briddon et al., 2008) are symptom-modulating, single-stranded DNA satellites that require the helper begomovirus for replication, spread in plant tissues, and plant-to-plant transmission by the whitefly vector of begomoviruses (*B. tabaci*; reviewed by Briddon and Stanley, 2006).

In addition to the begomovirus and betasatellites, a third group of ssDNA molecules, collectively known as alphasatellites (previously known as DNA 1; Briddon et al., 2009) have been shown to be associated with begomovirus-betasatellite complexes (Mansoor et al., 1999a, 2001; Saunders and Stanley, 1999). These satellite-like molecules encode a single product, a Rep protein with similarity to the Reps of nanoviruses; another family of plant-infecting single-stranded DNA viruses. Consequently, these alphasatellites are capable of autonomous replication in the cells of host plants, but they required the helper begomovirus for spread in plants and insect transmission (reviewed by Briddon and Stanley, 2006).

The family *Nanoviridae* consists of aphid transmitted viruses with isometric virions, approximately 18-20 nm in size, and is divided into two genera. The genus *Babuvirus*, at this time, contains only a single member BBTV. However, two further species have been proposed, Abacá bunchy top virus, and, Cardamom bushy dwarf virus (syn. Cardamom clump virus), and are likely to be ratified by the ICTV in the near future. The genus *Nanovirus* contains *Subterranean clover stunt virus* (SCSV) (Boevink et al., 1995), *Faba bean necrotic yellows virus* (FBNYV) (Katul et al., 1997; Katul et al., 1998) and *Milk vetch dwarf virus* (MDV; Sano et al., 1998). Even though, for each species, only a single molecule encodes the Rep that is responsible for trans-replication of the non-Rep encoding components (Timchenko et al., 1999; 2000) many nanovirus infections contain additional Rep-encoding components (Horser et al., 2001). The remaining Rep-encoding components, which depend upon
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

the helper nanovirus for systemic movement within plants and insect transmission, can be seen as satellite-like molecules. Whether these satellite-like components have any function in the pathogenicity of the virus has still to be established.

The satellite-like DNAs associated with begomoviruses have several similarities to the additional, satellite-like, Rep-encoding components of nanoviruses, including that they consist of single-stranded DNA, encode a rolling circle replication initiator protein (meaning that they are capable of autonomous replication in the cells of host plants) and have their predicted Rep promoters 3’ of the stem-loop structure that contains the nonanucleotide sequence (Horser et al., 2001; Briddon et al., 2004; 2009).

However, there are also some significant differences between the begomovirus and nanovirus-associated, satellite-like molecules. The begomovirus-associated DNAs are typically 200-300 nts larger. Much of this size difference is attributed to the presence, in the begomovirus-associated DNAs, of a region of sequence rich in adenine. This size increase is believed to be necessary to raise the size of a nanovirus DNA, from which begomovirus-associated alphasatellites molecules are believed to have arisen, and make it suitable for association with a begomovirus. This is necessary since there is a strict selection for unit (~2800 nts) and half-unit (~1400 nts) geminivirus DNA size for efficient movement and encapsidation. The Rep gene of begomovirus associated alphasatellites invariably encodes a 315 amino acid protein. Only a single begomovirus alphasatellite has been identified that encodes a Rep predicted to consist of 295 amino acids (Saunders, Bedford, and Stanley, 2002). The nanovirus-associated satellites typically encode a Rep consisting of 281-284 amino acids, whereas the genomic Rep-encoding component invariably encodes a 286 amino acid "master"Rep.

Bipartite begomoviruses are native to new world; their genomes consisting of two components, the first (known as DNA A) encoding all viral factors required for DNA replication, control of gene expression and encapsidation/insect transmission with the second (known as DNA B) encoding factors required for viral movement in host plants. A single monopartite begomovirus, *Tomato yellow leaf curl virus*, occurs
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

in the Caribbean and southern North America, having been introduced from the Old World during the 1990s.

Prior to 1999 the evidence suggested that the majority of begomoviruses occurring in the Old World also were bipartite, although a significant number of viruses lacking the DNA B component had been identified. The change in our understanding of the genetic make-up of begomoviruses came in 1999-2000, triggered by the identification of satellites of begomoviruses (Mansoor et al., 1999; Briddon et al., 2001; Saunders et al., 2000). The distribution of virus types now shows the monopartite begomoviruses to outnumber the bipartite viruses in the Old World, the majority of the monopartite viruses having an association with betasatellites.

Cotton leaf curl disease (CLCuD) is caused by a begomovirus-betasatellite complex consisting of multiple distinct begomoviruses supported by a single type of betasatellite (Mansoor et al., 2003b). The disease first appeared in the mid-1980s and was epidemic during the 1990s causing major losses to cotton production across Pakistan and western India. The introduction of resistant cotton varieties, obtained by conventional breeding and selection (Rahman et al., 2005), in the late 1990s restored production levels to their pre-epidemic levels. However, in 2002 a resistance breaking strain of the disease appeared and rapidly spread (Mansoor et al., 2003a; Mansoor, Zafar, and Briddon, 2006). Subsequently the resistance breaking strain (now known as the “Burewala” strain) was shown to be associated with a recombinant betasatellite (Amin et al., 2006) and a recombinant helper begomovirus, Cotton leaf curl Burewala virus, consisting of sequences derived from begomoviruses associated with earlier epidemic (L. Amrao, manuscript in preparation). At this time there are only tolerant varieties of cotton available to farmers and losses are again high (Rahman and Zafar, 2007). It was previously shown that, with the exception of Honey suckle yellow vein virus and Eupatorium yellow vein virus (Briddon et al., 2003), association of an alphasatellite with begomovirus-betasatellite complexes is the norm. For all Cotton leaf curl disease (CLCuD) complex affected plants from Pakistan and western India collected in the 1990s and early 2000s, which were examined for the presence of an alphasatellite, the presence of this molecule was shown.
3.2 Methodology

3.2.1 Virus source and DNA extraction

A total of 30 leaf samples originating from Sindh and Punjab provinces (Fig. 3.1) were collected in 2005 and 2006 from cotton plants with definite symptoms of CLCuD. In addition, a variety of other crops and weeds showing symptoms typical of begomoviruses were collected. The geographical origins and symptoms displayed by these plants are summarized in Table 3.1. Following collection, leaf samples were stored on ice during transport and then frozen at -70°C. DNA was extracted from all samples by the method of Doyle and Doyle (1990).

Figure 3.1 Symptoms exhibited by the crop and weed plants used for the alphasatellite study. The species shown are Conyza stricta (A), Xanthium strumarium (B), Sonchus arvensis (C), mungbean (Vigna radiata, D), chickpea (Cicer arietinum, E), cucumber (Cucumis sativus, F), chili (Capsicum annuum, G), tomato (Solanum Lycopersicum, H) and cotton (Gossypium hirsutum, I).
Figure 3.2 Map of Pakistan showing the geographical origin of the plant samples used in the alphasatellite study. ★ Major cities  ● Sampling areas
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

### Table 3.1 Detection of alphasatellites in crops and weeds

<table>
<thead>
<tr>
<th>Host/plant</th>
<th>Scientific name</th>
<th>Begomovirus species*</th>
<th>Alphasatellite present</th>
<th>Clone name</th>
<th>Year</th>
<th>Location</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cotton</td>
<td><em>Gossypium hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Faisalabad</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuShV</td>
<td>Yes</td>
<td>DSB</td>
<td>2005</td>
<td>Shadadpur</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2006</td>
<td>Faisalabad</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuSiV</td>
<td>Yes</td>
<td>DSC</td>
<td>2006</td>
<td>Shadadpur</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Arifwalla</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Khenewal</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Layyah</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Muzzafargarh</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Multan</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Rahim Yar Khan</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Vehari</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Sualiwal</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Cotton</td>
<td><em>G. hirsutum</em></td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2005</td>
<td>Faisalabad</td>
<td>LC,E,VD,VS,LE</td>
</tr>
<tr>
<td>Weed</td>
<td>Conyza stricta</td>
<td>CLCuBV</td>
<td>Yes</td>
<td>WD15</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Conyza stricta</td>
<td>CLCuBV</td>
<td>Yes</td>
<td>WD17</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Conyza stricta</td>
<td>CLCuBV</td>
<td>Yes</td>
<td>WD18</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Xanthium strumarium</td>
<td>ToLCGV</td>
<td>Yes</td>
<td>CL2</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Sonchus arvensis</td>
<td>AlYVV</td>
<td>Yes</td>
<td>UK5</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Sonchus arvensis</td>
<td>AlYVV</td>
<td>Yes</td>
<td>UK7</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Sonchus arvensis</td>
<td>AlYVV</td>
<td>Yes</td>
<td>UK8</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Sonchus arvensis</td>
<td>AlYVV</td>
<td>Yes</td>
<td>UK12</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Weed</td>
<td>Sonchus arvensis</td>
<td>AlYVV</td>
<td>Yes</td>
<td>UKV</td>
<td>2007</td>
<td>Samundri</td>
<td>VY,VT,LC</td>
</tr>
<tr>
<td>Tomato</td>
<td>Solanum Lycopersicum</td>
<td>CLCuBV</td>
<td>No</td>
<td>-</td>
<td>2007</td>
<td>Rahim Yar Khan</td>
<td>LC,VY,VS</td>
</tr>
<tr>
<td>Tomato</td>
<td>Solanum Lycopersicum</td>
<td>CLCuRV</td>
<td>No</td>
<td>-</td>
<td>2007</td>
<td>Faisalabad</td>
<td>LC,VY,VS</td>
</tr>
<tr>
<td>Tomato</td>
<td>Solanum Lycopersicum</td>
<td>ToLCNDV</td>
<td>No</td>
<td>-</td>
<td>2008</td>
<td>Sualiwal</td>
<td>LC,VY,VS</td>
</tr>
<tr>
<td>Chilies</td>
<td>Capsicum annuum</td>
<td>ChiLCuV</td>
<td>No</td>
<td>-</td>
<td>2008</td>
<td>Sualiwal</td>
<td>LS,LC,Y</td>
</tr>
<tr>
<td>Soyabeans</td>
<td>Glycine max</td>
<td>MYMIV</td>
<td>No</td>
<td>-</td>
<td>2008</td>
<td>Nawabshah</td>
<td>LC,VY</td>
</tr>
<tr>
<td>Chickpea</td>
<td>Cicer arietinum</td>
<td>CpCDPKV</td>
<td>No</td>
<td>-</td>
<td>2006</td>
<td>Mianwali</td>
<td>Y,VS,LC</td>
</tr>
<tr>
<td>Mungbeans</td>
<td>Vigna radiate</td>
<td>MYMIV</td>
<td>No</td>
<td>-</td>
<td>2006</td>
<td>Layyah</td>
<td>LC,VD</td>
</tr>
<tr>
<td>Cucumber</td>
<td>Cucumis sativus</td>
<td>SLCV</td>
<td>No</td>
<td>-</td>
<td>2008</td>
<td>Sualiwal</td>
<td>LD,LC,VS</td>
</tr>
</tbody>
</table>

* The virus species in each case was identified by sequencing the entire genome in a separate study.
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

Figure 3.3  Southern blots of DNA extracts from CLCuD field infected cotton plants (A) and φ29 polymerase amplification products from DNA extracts of CLCuD affected cotton plants digested with PstI (B). Blots were probed with a radioactive P32 labelled alphasatellite probe with CLCuMA-[PK:Fai1:98] AJ132344. For panel A the gel was loaded with samples collected from Sindh province (Nawabshah and Mirpur Khaus districts; lane 1-6) in 2005 and Punjab province (Faisalabad district; lanes 7-12) in 2006. Approx. 20 μg of DNA was loaded in each lane with the exception of lane M which was loaded with approx. 0.1 μg of an undigested clone of an alphasatellite (CLCuMuA-[PK:Fai1:98]) as a hybridization control. For gel B the samples included the two Φ29 polymerase amplification products of reactions containing DNA extracts which were shown to contain alphasatellite (among the samples from 2006 lane 5 show the presence of alphasatellite while the lane 5 from 2005 also show a sharp band). Approx. 5 μg of Φ29 polymerase amplification product was loaded in each case.
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

3.3 Results

3.3.1 Detection of alphasatellites in crops and weeds

Screening of 30 nucleic acid extracts from cotton, 6 of Conyza stricta, 9 of Sonchus arvensis, 5 of Xanthium strumarium, 13 of cucumber, 25 of chickpea, 15 of soyabean, 18 of chillies and 27 of tomato with universal alphasatellite primers (primer pair DNA101/DNA102; Bull et al., 2003) identified only two from cotton in which the satellite-like molecule could be amplified by PCR and nine from three different weed hosts, three from Conyza stricta (family Asteraceae), five from Sonchus arvensis (family Asteraceae) and one from Xanthium strumarium (family Asteraceae), yielding the expected amplification product of approx. 1370 nucleotides (results not shown). All the remaining weed and other crop samples were negative for alphasatellites. Selected amplification products were cloned into pTZ57R/T using the InsT/Aclone kit (Fermentas).

Similarly the analysis of 30 DNA extracts from affected cotton plants identified only two which gave a positive hybridization signal to the CLCuD alphasatellite probe when washed at intermediate stringency. Due to the low levels of diversity between alphasatellites from distinct sources, hybridization at this stringency would be expected to detect the presence of the molecule. To account for the possibility that alphasatellites could be present in many plants but at levels below the detection threshold of PCR and Southern blot hybridization, rolling circle amplification (RCA (Haible et al., 2006) using Φ29 polymerase (Fermentas) to initially amplify all circular DNA molecules from the ten randomly selected DNA extracts of CLCuD affected plants. A high molecular weight product was produced in all RCA reactions. Equal amounts of the amplification product were then digested with the restriction endonuclease PsrI, run on an agarose gel, transferred to nitrocellulose and probed by Southern hybridization for the presence of alphasatellite (Fig.3.4). Only for the two samples, for which PCR had shown the presence of alphasatellite, was hybridization to RCA product positive. The presence of alphasatellite in only a small number of plants indicates that, in contrast to earlier analyses, alphasatellites are no longer a constant feature of the CLCuD begomovirus-DNA β complex.
3.3.2 Sequencing and Sequence Analysis.

3.3.2.1 Analysis of alphasatellites isolated from cotton

The complete sequences of alphasatellite clones obtained from cotton and weeds were determined. These sequences are available in the databases under the accession numbers shown in (Table 3.2). The features of the alphasatellites are shown in Table 3.2. They have the typical arrangement of alphasatellites, consisting of a single open reading frame, a region rich in adenine (A-rich) and a predicted hairpin structure, containing the stem loop sequence TAGTATTAC, with similarity to the origin of virion-strand DNA replication of the geminiviruses and nanoviruses (Briddon et al., 2004; Mansoor et al., 1999b; Saunders and Stanley, 1999).

The stem sequences of the two molecules from cotton are distinct, being GCTCCGCCC and TGGCTCCGACC for the 2005 and 2006 isolates, respectively. Nevertheless, they fall into the groups of stem sequences (groups 1 and 2, respectively) shown previously to be present in alphasatellites identified from CLCuD affected cotton on the subcontinent (Briddon et al., 2004). Sequence comparisons with all available alphasatellites in the databases show [PK:DSB:05] to have the highest nucleotide sequence identity (82.8%) with an alphasatellite isolated from TLCD affected tomato in Pakistan (ToLCA-[PK:NOB2:97]). [PK:DSC:06] showed the highest levels of nucleotide sequence identity (80.1%) to an alphasatellite isolated from CLCuD affected cotton CLCuMuA-[PK:Fai1:98], the first such molecule to be identified (Mansoor et al., 1999b).

Since 83% sequence identity has recently been proposed as species demarcation threshold for distinct alphasatellites (Briddon et al., 2009), the alphasatellites associated with cotton, are isolates of a new species for which the names Cotton leaf curl Sindh alphasatellite (CLCuSiA) and Cotton leaf curl Shadadpur alphasatellite (CLCuShA), respectively, are proposed.
3.3.2.2 Analysis of alphasatellites isolated from weeds

The alphasatellites isolated from *Conyza stricta* showed <73% sequence identity to Tomato yellow leaf curl China alphasatellite (TYLCCNA) and <74% identity to Tobacco curly shoot alphasatellite TbCSA isolates (Tab. 3.3). The highest level of sequence identity (72.8%) was between [PK:Wd18:Con:08] and Tomato yellow leaf curl China alphasatellite TYLCCNA-[CN:Yn89:02]AJ579358. Based on the 83% species demarcation threshold, these satellites are isolates of a newly identified alphasatellite species for the name Conyza yellow leaf curl alphasatellite (ConYLCA) is proposed.

Out of the three alphasatellites isolated from *Conyza stricta* one, [PK:WD15:Con:08], has a mutated Rep, predicted to consist of 214 amino acids. The mutation occurred at position 712, where a deleted G introduces a premature stop codon. For the other two sequences, the predicted Rep is 314aa, where as the Rep proteins of alphasatellites are typically 315aa. This is due to the deletion of a threonine residue at amino acid position 16. Although this is atypical of the vast majority of alphasatellites, an identical deletion was recently found in a clone of an alphasatellite isolated from potato originating from Lahore (Pakistan). This was identified as a new alphasatellite species, for which the name Potato leaf curl alphasatellite was proposed (Mubin et al., 2009). Whether this single amino acid change has any effect on the function/activity of the Rep protein has not been determined.

A total of six alphasatellite clones were obtained from *X. strumarium* and *S. arvensis* (Table 3.2). These are typical of all previously isolated alphasatellites and encode a Rep protein predicted to consist of 315aa. They show between 85.5 and 95.6% nucleotide sequence identity to isolates of Hibiscus leaf curl alphasatellite (HLCuA; Table 3.3). The highest levels of identity (95.6%) were between [PK:UKV:Son:08] and HLCuA-[PK:F22;00]AJ512952. This indicates that all these clones represent isolates of HLCuA, being above the proposed species demarcation threshold of 83%.
3.3.3 Phylogenetic analysis of alphasatellite sequences

A phylogenetic tree based upon an alignment of the alphasatellite sequences produced here with selected alphasatellite sequences from the databases is shown in Fig 3.5. This shows the alphasatellites isolated from S. arvensis and X. strumarium to closely associate with isolates of Hibiscus leaf curl alphasatellite (HLCuA), confirming their identification as isolates of HLCuA. This alphasatellite species was first identified by Briddon et al. (2003) and occurs widely across southern Asia, having been identified in Pakistan, India and China. Surprisingly this species has also been identified in Egypt (Briddon et al., 2003). The short branch lengths show all isolates of HLCuA from Pakistan to be closely related. However, these are distinct from HLCuA isolates from India, which group with and are closely related to isolates from China, although they form a separate group. This thus suggests that HLCuA in these three countries are evolving independently but have only diverged relatively recently.

The newly identified alphasatellite species from C. stricta, ConYLCA, groups with TYLCCNA, but is distinct from this, as evidenced by the long branch lengths. Only a single isolate of TYLCCNA is available in the databases at this time and originates from China.

The alphasatellites isolated from cotton collected in Sindh each group with alphasatellites which have previously been identified in cotton or which is connected to cotton in some other way. CLCuSiA clusters with CLCuMA, the first alphasatellite species identified that was isolated from cotton affected by the Multan strain of CLCuD (Mansoor et al., 2009). In contrast, CLCuShA clusters with ToLCA. Although ToLCA has not been shown to infect cotton, the present CLCuMB (that associated with the “Burewala” strain of CLCuD) is a recombinant molecule consisting for the most part of sequences from the “Multan” strain CLCuMB and a small amount of sequence, within the satellite conserved region, originating from a ToLCB (Amin et al., 2006). Similarly a begomovirus first identified in tomato, Tomato leaf curl Bangalore virus (ToLCBV-B[IN:Fat:Cot] AY456684), was isolated from cotton originating from India (Kirthi et al., 2004). There is thus a definite link between tomato and cotton with respect to the components of the CLCuD complex and it may not be surprising to find a tomato-related alphasatellite present in cotton. Tomato may act as an alternate host for the CLCuD-components or these components are found in tomato due to the polyphagous nature of the whitefly vector, B. tabaci. Alphasatellites appear to have few host range constraints and appear able to associate with many begomoviruses and even
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

curtoviruses (Briddon et al., 2003; Saunders et al., 2002). The alphasatellites isolated from cotton collected in Sindh each group with alphasatellites which have previously been identified in cotton or which is connected to cotton in some other way. CLCuSiA clusters with CLCuMA, the first alphasatellite species identified that was isolated from cotton affected by the Multan strain of CLCuD (Mansoor et al., 2009). In contrast, CLCuShA clusters with ToLCA. Although ToLCA has not been shown to infect cotton, the present CLCuMB (that associated with the “Burewala” strain of CLCuD) is a recombinant molecule consisting for the most part of sequences from the “Multan” strain CLCuMB and a small amount of sequence, within the satellite conserved region, originating from a ToLCB (Amin et al., 2006). Similarly a begomovirus first identified in tomato, Tomato leaf curl Bangalore virus (ToLCBV-B[IN:Fat:Cot] AY456684), was isolated from cotton originating from india (Kirthi et al., 2004). There is thus a definite link between tomato and cotton with respect to the components of the CLCuD complex and it may not be surprising to find a tomato-related alphasatellite present in cotton. Tomato may act as an alternate host for the CLCuD-components or these components are found in tomato due to the polyphagous nature of the whitefly vector, B. tabaci. Alphasatellites appear to have few host range constraints and appear able to associate with many begomoviruses and even curtoviruses (Briddon et al., 2003; Saunders et al., 2002).
### Table 3.2 Features of alphasatellite molecules amplified from cotton and weed hosts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[PK:WD17:Con: 08]</td>
<td><em>C. stricta</em></td>
<td>AM749493</td>
<td>1369</td>
<td>74-1018/945</td>
<td>314</td>
<td>39.02</td>
</tr>
<tr>
<td>[PK:WD18:Con: 08]</td>
<td><em>C. stricta</em></td>
<td>AM749494</td>
<td>1369</td>
<td>74-1018/945</td>
<td>314</td>
<td>39.02</td>
</tr>
<tr>
<td>[PK:CL2:Xan: 08]</td>
<td><em>Xanthium strumarium</em></td>
<td>AM930247</td>
<td>1375</td>
<td>78-1025/948</td>
<td>314</td>
<td>36.54</td>
</tr>
<tr>
<td>[PK:UK5:Son: 08]</td>
<td><em>Sonchus arvensis</em></td>
<td>AM884370</td>
<td>1375</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.54</td>
</tr>
<tr>
<td>[PK:UKV5:Son: 08]</td>
<td><em>S. arvensis</em></td>
<td>AM930244</td>
<td>1374</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.54</td>
</tr>
<tr>
<td>[PK:UK12:Son: 08]</td>
<td><em>S. arvensis</em></td>
<td>AM930245</td>
<td>1375</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.54</td>
</tr>
<tr>
<td>[PK:UK7:Son : 08]</td>
<td><em>S. arvensis</em></td>
<td>AM930246</td>
<td>1375</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.54</td>
</tr>
<tr>
<td>[PK:UK8:Son: 08]</td>
<td><em>S. arvensis</em></td>
<td>AM930248</td>
<td>1374</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.54</td>
</tr>
<tr>
<td>[PK:DSB:05]</td>
<td><em>Gossypium hirsutum</em></td>
<td>AM711115</td>
<td>1380</td>
<td>79-1026/948</td>
<td>315</td>
<td>38.97</td>
</tr>
<tr>
<td>[PK:DSC:06]</td>
<td><em>G. hirsutum</em></td>
<td>AM711116</td>
<td>1374</td>
<td>77-1024/948</td>
<td>315</td>
<td>39.45</td>
</tr>
</tbody>
</table>
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

Figure 3.4 Neighbour joining Phylogenetic dendrogram based upon an alignment of the complete nucleotide sequences of alphasatellite isolates from Punjab and Sindh with selected alphasatellites from the database. Numbers at the nodes are percentage bootstrap confidence values (1000 replicates). Vertical lines are arbitrary and horizontal lines are proportional to calculated mutation distances. The tree is rooted on the betasatellite CLCuMB-PK:Fai:Tom:05](AM490309). The species used were Ageratum yellow vein India alphasatellite (AYVIA), Ageratum yellow vein Pakistan alphasatellite (AYVPKA), Ageratum yellow vein Kenya alphasatellite (AYVKA), Cotton leaf curl Multan alphasatellite (CLCuMA), Hibiscus leaf curl alphasatellite (HLCuA), Malvestrum yellow mosaic alphasatellite (MalYMA), Malvestrum yellow mosaic Hainan alphasatellite (MalYMHnA), Okra leaf curl alphasatellite (OLCuA), Sida yellow vein Vietnam alphasatellite (SiYVVA), Tobacco curly shoot alphasatellite (TbCSA), Tomato leaf curl alphasatellite (ToLCA), Tomato yellow leaf curl China alphasatellite (TYLCCNA), Tomato leaf curl China Yunan alphasatellite (TYLCYnA). The alphasatellite isolate descriptors are as given in (Briddon et al., 2009). The sequences produced as part of this study are highlighted.
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

Table 3.3  The highest and the lowest percentage nucleotide sequence identities of the alphasatellites isolated from cotton and weeds with all available sequences of other alphasatellite available in the databases

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YLCCNA (6)</td>
<td>100</td>
<td>67.4</td>
<td>78.3</td>
<td>78.3</td>
<td>78.5</td>
<td>78.4</td>
<td>77.8</td>
<td>78.3</td>
<td>66.4</td>
<td>66.4</td>
<td>65.2</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>TkCSA (6)</td>
<td>100</td>
<td>64.6</td>
<td>68.4</td>
<td>66.9-69.2</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td></td>
</tr>
<tr>
<td>HLCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>TglCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>GtLCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>MtlYMA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>SiLCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>MinLCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>OLCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
<tr>
<td>OKLCuA (4)</td>
<td>100</td>
<td>68.8</td>
<td>68.9</td>
<td>68.9</td>
<td>69.3</td>
<td>68.6</td>
<td>69.0</td>
<td>66.9</td>
<td>62.3-70.4</td>
<td>74.4-78.8</td>
<td>75.5-82.8</td>
<td>69.4-71.4</td>
<td>66.4-65.5</td>
<td>63.9-68.1</td>
<td>67.3</td>
<td>76.0</td>
<td>65.3</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
<td>[PK:DSR:05]</td>
</tr>
</tbody>
</table>

* The number of sequences available which were used in the comparisons.
3.4 Discussion

Since the identification of the first betasatellite in 2000 (Saunders et al., 2000) there has been a flurry of interest in begomovirus/betasatellite complexes. The wealth of information that has been generated, including the sequences of numerous new begomovirus species and the complete sequences of over 260 betasatellite isolates (Briddon et al., 2008), indicates the importance these virus-satellite complexes have to agriculture across the Old World.

Despite the recent interest in betasatellites, the alphasatellites have not come in for the same level of interest; only 68 sequences were available in the databases as of December 2007 (Briddon et al., 2009). Likely this lack of interest is due to the findings of Mansoor et al. (1999) and Saunders et al. (1999) who reported that alphasatellites appear to play no discernable role in pathogenicity, possibly acting as a minor “dampener” of severity of the disease with which they are associated. Nevertheless, the fact that with all isolates of the “Multan” strain of CLCuD examined, as well as many other begomovirus-betasatellite complexes, alphasatellites were always present (Briddon et al., 2004; R.W. Briddon personal communication) suggested that they must play some important role, even if this satellite-like component only provides a very subtle selective advantage to the virus.

Although in all earlier samples of CLCuD affected cotton that were examined an alphasatellite was present, the data presented here show that this is no longer the case. For none of the samples of CLCuD affected cotton collected from the Punjab was an alphasatellite present. The CLCuD complex has gone through a major change since the alphasatellites were first identified. The previous strain of the disease, now known as the “Multan strain”, has been almost entirely displaced by the “Burewala strain” (Amin et al., 2006; Amrao et al., 2009). This new strain is resistance breaking (Mansoor et al., 2003), able to infect cotton varieties derived from the germplasm LRA5166 and CP15/2 with resistance to the “Multan” strain (Rahman et al., 2005).

The Burewala strain is associated with a recombinant betasatellite, consisting of, for the most part, sequences derived from the Multan strain Cotton leaf curl Multan betasatellite (CLCuMB) but with a small region of sequence within the satellite conserved region (SCR) derived from a Tomato leaf curl betasatellite (Briddon et al., 2003; Amin et al., 2006).
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

The significance of this recombination between a cotton and a tomato betasatellite remains unclear, in a large part due to the fact that we have yet to understand the precise function of the SCR of betasatellites. Saunders et al. (2008) have recently shown it to possibly play a major part in the transreplication of the betasatellite by the helper virus-encoded Rep. Even more recently the begomovirus associated with the “Burewala” strain has been sequenced and shown to also be recombinant, consisting of sequences derived from two viruses associated with the Multan strain of the disease (Amrao et al., 2009).

The absence of an alphasatellite is thus just one of several changes that have occurred in the CLCuD complex of southern Asia in recent years. Upon their first identification it was suggested that the function of alphasatellite may be to down-regulate the severity of the symptoms caused by the complex, by mopping up cellular resources (Mansoor et al., 2000; Saunders and Stanley, 1999). Some evidence for this has been provided. Wu and Zhou (2005) demonstrated that the presence of alphasatellite reduced virus DNA levels, but only in the absence of the betasatellite, although symptoms in both cases were attenuated at the early stages of infection. More recently, Hussain et al. (2009) have shown that the alphasatellite associated with the Multan strain of CLCuD (CLCuMA) is the target of RNA silencing mediated host defense (siRNAs complimentary to the alphasatellite are produced in plants). They suggested that possibly the alphasatellite functions to “deflect” the host defense response in a futile response to the alphasatellite, thus diluting the response to the more important virus and satellite components. However, these results were not too convincing. It is thus clear that further investigation of the function(s) of alphasatellites in begomovirus-betasatellite infections is required.

If alphasatellite acts as a “dampener” of the virus infections, then we must assume that its dampening effect is no longer a requirement of the CLCuD complex presently in the field in Pakistan. Possibly a more pathogenic complex is required (thus one lacking a “dampener”) to achieve infection of resistant cotton varieties. Alternatively, other factors may complement the function of the missing alphasatellite. Defective interfering (di)DNAs are a common feature of many begomovirus infections. These are defective (half or quarter unit length) geminivirus genomic molecules, usually derived preferentially from the DNA B component for bipartite begomoviruses, which can reduce the severity of infections by interfering with virus replication; presumably by mopping-up cellular resources required for virus replication (Patil et al., 2006). In many respects the diDNAs have properties similar to
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

satellites, particularly alphasatellites, in reducing virus DNA levels in plants. The presence of diDNAs in CLCuD affected plants has not been extensively investigated, although such molecules have been identified (Briddon et al. 2001; Nouman Tahir and Sohail Akhtar, personal communication). Certainly the possibility that diDNAs play a part in CLCuD, and their potential role in complementing the absence of an alphasatellite, should be investigated further.

In contrast to the Punjab, alphasatellites were identified in CLCuD affected cotton originating from Sindh. This area was not affected by the Multan strain epidemic of the 1990s and only recently has the first report of CLCuD occurring in Sindh been made (Mansoor et al., 2006). The absence of CLCuD in this area during the 1990s was attributed to the presence of a distinct B. tabaci biotype which, it was argued, may possibly not efficiently transmit the viruses causing CLCuD or that does not have a host range extending to cotton (Simón et al., 2003). For this reason CLCuD resistant cotton varieties have not been used in Sindh and the Burewala strain of the disease does not occur there (Luqman Amrao, personal communication).

An analysis of a small number of other crop plant species with symptoms typical of begomovirus infection was also unable to identify the presence of alphasatellite. Although mungbean and soybean are typically infected by bipartite begomoviruses, and these have not previously been shown to associate with alphasatellite, chili and tomato are commonly affected by monopartite begomovirus-betasatellite complexes and were previously shown to harbour alphasatellite (Briddon et al., 2003). Alphasatellites were readily isolated from a number of weeds. Thus there appears to be a difference between crops and weeds with respect to the presence of alphasatellites. In view of the low numbers of samples analysed, this is far from conclusive, but it is evident that the situation now differs from that reported earlier. Alphasatellites are no longer a constant feature of begomovirus-betasatellite complexes in Pakistan. The precise reason for this remains unclear.

The alphasatellites identified in Sindh are distinct from those shown earlier to be associated with CLCuD and represent newly identified alphasatellite species. However, both are closely related to previously identified species that have an association with cotton. CLCuSiA is closely related to CLCuMA, the first alphasatellite identified (Mansoor et al., 1999b). CLCuShA is most closely related to ToLCA (Briddon et al., 2003). There is a close
An Alphasatellite is no Longer a Constant Feature of the Cotton Leaf Curl Disease Begomovirus/Betasatellite Complex

association of tomato and cotton with respect to begomoviruses. The CLCuMB strain presently in the field is a recombinant between CLCuMB (Multan strain) and a ToLCB. At least one virus first identified in tomato, Tomato leaf curl Bangalore virus (Kirthi et al., 2004), was identified in cotton affected by CLCuD. It is likely that, due to the movement of the vector between plant species, that tomato acts as a reservoir for the components of the CLCuD complex and may be the species in which recombination occurs. It is thus not surprising to find an alphasatellite species closely related to a typical tomato alphasatellite in cotton. The identification of three new species here indicates that we have yet to identify the complete diversity of these molecules and it is likely that a more thorough examination will uncover further diversity in the future.
Chapter 4
4.1 Introduction

Gene expression by most eukaryotes is regulated to some extent by transcriptional and posttranscriptional silencing mediated by siRNAs produced from double-stranded templates by enzymes of the Dicer family (Margis et al., 2006). Gene silencing based on RNAi was first reported to be induced by the introduction of dsRNA in *Caenorhabditis elegans* cells in 1998 (Fire et al., 1998). dsRNA-dependent RNA silencing is a quite common phenomenon in plants (Sato, 2005). It is realized that such co-suppression or some virus-resistance phenomenon in plants can be based on RNA-degradation via dsRNA (Baulcombe, 2004). In fact, such mRNA degradation based on dsRNA is a common genetic mechanism for controlling gene expression in both plants and animals (Baulcombe, 2004; Meister and Tuschl, 2004). dsRNA molecules are generated through aberrant gene expression, virus infection or a tandem repeat sequence due to the insertion of a transposon. The resulting dsRNA molecules are digested into 21–25 nucleotide-long siRNA by Dicer (an RNaseIII-like enzyme). This siRNA acts as a template for the targeted degradation of mRNA by RISC (RNA-induced silencing complex). So far, many molecular components in the RNAi machinery have been characterized and secondary RNAi based on RNA-dependent RNA polymerase (RdRp) has been reported in some animals including *C. elegans*. RNAi has great potential as a tool for down-regulating gene expression. RNAi induced by the expression of dsRNA in target cells induced much stronger and selective gene silencing. Transient RNAi has also been successfully used for the comprehensive analysis of gene function (Sato, 2005). The identification of 21–25 nucleotide-long siRNA, as the active ingredient for RNAi machinery, has enabled it to be used for RNAi in mammalian cells and expanded the scope of RNAi (Elbashir et al., 2001).

In 1990, the first reports on gene silencing induced by co-suppression were published (Napoli, 1990; van der Krol, 1990) when researchers were attempting to enhance the purple pigmentation of petunia petals by over-expression of *CHALCONE SYNTHASE (CHS)*, a gene in to the anthocyanin pathway. But this approach induced multicolored and white petals, caused by silencing of the endogenous chalcone synthase gene. Unfortunately, the white color of the petals was observed in next
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

generation, occasionally reversion of the plant from white to purple was observed, demonstrating that the occurrence did not involve permanent DNA modification.

Virus based vectors carrying host-derived sequence inserts referred to as virus-induced gene silencing (VIGS) vectors. VIGS has been developed and used to induce silencing of the corresponding genes in infected plants. It can be targeted, in a sequence specific manner, against either endogenous or transgene mRNA to generate mutant phenotypes for assigning function to unknown genes (Baulcombe, 1999). It is based on a silencing mechanism that regulates gene expression by the specific degradation of RNA (Cogoni, 1997). RNA silencing is a conserved phenomenon that has been reported in animals, plants and fungi (Cogoni, 1997), and involves complex molecular machinery (Lu et al., 2003). In plants, at least three pathways with several biological roles have been reported; all three coincide in the cleavage of double-stranded RNA (dsRNA) into short (21–25 nt) RNA molecules (Baulcombe, 2004). One of the roles of RNA silencing is to defend plants against viruses. Thus, when a virus carrying a fragment of an endogenous gene (or a transgene) is inoculated into a plant, the silencing mechanism is triggered and a specific RNA degradation process is turned on to destroy all transcripts derived from the modified virus (Lu et al., 2003). In addition to local RNA degradation, a mobile silencing signal is produced that brings the instructions for specific degradation far away from the inoculation point (Muangsan et al., 2004).

Examples of the application of this technology include the use of a Potato virus X (PVX)-based VIGS vector to down-regulate nitrate reductase (NR) gene in N. benthamiana leaves. The expression of Nb14-3-3a as well as Nb14-3-b genes in the leaves of PVX-14-3a-infected plants was repressed altogether. The binding of 14-3-3 proteins to phosphorylated NR leads to substantial decrease in NR activity of leaves in the dark. This confirmed that Nb14-3-3a and/or Nb14-3-3b proteins are indeed involved in the inactivation of NR activity in the dark in N. benthamiana (Hirano et al., 2007). The PVX vector has also been used to silence the putative cellulose synthase genes (CesA) in N. benthamiana using the VIGS approach (Burton, 2000). Similarly a TRV-based vector was used to investigate the phenylpropanoid volatile production pathway in petunia. This high-throughput reverse-genetics approach was applicable to both regulatory and structural genes responsible for volatile production
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

and is expected to be useful for large scale scanning and functional characterization of novel scent genes (Spitzer, 2007).

VIGS is useful for the knockout of those genes in which the mutation rate is very high. VIGS is quicker and easier than other techniques, since it does not rely on the production of stable transformants. It can also be useful to silence more than one gene at the same time in single experiment (Carrillo-Tripp, Shimada-Beltran, and Rivera-Bustamante, 2006). VIGS can also be used to test the function of essential genes because the targeted seedlings have already germinated and produced leaves before silencing begins (Peele et al., 2001; Jordan, 2007). Combinations of genes can be down-regulated simultaneously from the same vector (Peele et al., 2001; Turnage, 2002), allowing transcripts of multiple gene family members to be down-regulated.

VIGS offers a great opportunity to test gene function through silencing, using homologous clone sequence. Only partial gene sequences are needed to initiate gene silencing, ranging from about 90-150 bp if inserted as part of a viral gene (Peele et al., 2001; Jordan, 2007), or up to approximately 400-800 bp if inserted as a gene replacement (Kjemtrup et al., 1998). Silencing typically begins within a few weeks of inoculation and can extend throughout the flowering period. VIGS can also be used to test the function of essential genes because the targeted seedlings have already germinated and produced leaves before silencing begins (Peele et al., 2001; Jordan, 2007). Combinations of genes can be down-regulated simultaneously from the same vector (Peele et al., 2001; Turnage, 2002), allowing transcripts of multiple gene family members to be eliminated. This ability could be extremely useful in cotton because of its allotetraploid genome.

However, virus induced gene silencing systems have some drawbacks too. Importantly VIGS can confuse phenotypic expression between the silenced and the non silenced parts of the plants. Sometimes VIGS does not show silencing readily and is not uniform throughout the whole plant. Also the level of silencing can vary within the plant and between experiments. For some VIGS systems, such as those that are based on PVX and TMV, the virus symptoms can mask the suppression phenotypes. As a result they cannot be used for the study of genes involved in leaf, flower, shoot and fruit development (Burch-Smith et al., 2004).
### Table 4.1 Geminiviruses and associated components used as VIGS vectors

<table>
<thead>
<tr>
<th>Virus*</th>
<th>Strategy**</th>
<th>Genes silenced</th>
<th>Hosts silenced</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACMV</td>
<td>CP</td>
<td>PDS, Su</td>
<td><em>Nicotiana benthamiana</em></td>
<td>(Fofana et al., 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Manihot esculenta</em></td>
<td></td>
</tr>
<tr>
<td>CaLCuV</td>
<td>NSP, CP</td>
<td>PDS, Su</td>
<td><em>Arabidopsis thaliana</em></td>
<td>(Muangsan et al., 2004; Turnage, 2002)</td>
</tr>
<tr>
<td>CLCrV</td>
<td>CP</td>
<td><em>PDS, CHL1</em></td>
<td><em>Gossypium hirsutum</em></td>
<td>(Tuttle et al., 2008)</td>
</tr>
<tr>
<td>PHYVV</td>
<td>CP</td>
<td>Su</td>
<td><em>Capsicum annum</em></td>
<td>(Carrillo et al., 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Nicotiana tabacum</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Solanum lycopersicum</em></td>
<td></td>
</tr>
<tr>
<td>TGMV</td>
<td>CP, NSP</td>
<td>Su, PCNA</td>
<td><em>N. benthamiana</em></td>
<td>(Kjemtrup et al., 1998; Peele et al., 2001)</td>
</tr>
<tr>
<td>TYLCCNB</td>
<td>C1</td>
<td>FROI</td>
<td><em>Solanum lycopersicum</em></td>
<td>(Tao et al., 2008)</td>
</tr>
</tbody>
</table>

* ACMV (African cassava mosaic virus), CaLCuV (Cabbage leaf curl virus), CLCrV (Cotton leaf crumple virus), PHYVV (Pepper huasteco yellow vein virus), TGMV (Tomato golden mosaic virus) and TYLCCNB (Tomato yellow leaf curl China betasatellite).

** CP (coat protein replacement), NSP (transcriptional fusion with the nuclear shuttle protein gene), C1 (betasatellite encoded βC1 replacement), Su (Sulphur), PCNA (Proliferating cell nuclear antigen), PDS (Phytoene desaturase), CHL (Chelatase) and FROI (ferric chelate reductase gene).
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

Table 4.2  Plant viruses that have been used as VIGS vectors

<table>
<thead>
<tr>
<th>Virus*</th>
<th>Target genes**</th>
<th>Silencing Host</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSV</td>
<td>RCY1</td>
<td><em>Arabidopsis thaliana</em></td>
<td>(Igarashi et al., 2009)</td>
</tr>
<tr>
<td>BPMV</td>
<td>Actin</td>
<td><em>Glycine max</em></td>
<td>(Zhang et al., 2009)</td>
</tr>
<tr>
<td>BSMV</td>
<td>P23k</td>
<td><em>Hordeum vulgare</em></td>
<td>(Ai Oikawa1, 2007)</td>
</tr>
<tr>
<td>PEBV</td>
<td>Nin</td>
<td><em>Pisum sativum</em></td>
<td>(Constantin et al., 2008)</td>
</tr>
<tr>
<td>PopMV</td>
<td>GFP</td>
<td><em>Nicotiana benthamiana</em></td>
<td>(Naylor et al., 2005)</td>
</tr>
<tr>
<td>PVX</td>
<td>PDS</td>
<td><em>N. benthamiana</em></td>
<td>(Kumagai, 1995)</td>
</tr>
<tr>
<td>TMV</td>
<td>TIR,NBS,LRR</td>
<td><em>N. benthamiana</em></td>
<td>(Liu et al., 2002)</td>
</tr>
<tr>
<td>TYMV</td>
<td><em>PDS</em></td>
<td><em>A. thaliana</em></td>
<td>(Stephanie et al., 2008)</td>
</tr>
</tbody>
</table>

* ALSV (Apple latent spherical virus), BPMV (Bean pod mottle virus) PEBV(Pea early browning virus), TYMV(Turnip yellow mosaic virus), BSMV(Barley stripe mosaic virus), TMV(Tobacco mosaic virus), PVX(Potato virus x) and (PopMV) Poplar mosaic virus.

** RCY1 (a resistant gene to an yellow strain of Cucumber mosaic virus in *A. thaliana*), ACTIN (soybean homologues of genes involved in plant defense, translation, and the cytoskeleton in shoots and in roots), TIR (Toll-interleukin-1 Receptor), NBS (nucleotide binding site), LRR (leucine-rich repeat), PDS (Phytoene desaturase), Nin (Nodule inception gene), P23k (monocot-unique protein wall formation in barley leaves) and GFP (green fluorescence protein).
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

The majority of the available VIGS vectors have relatively narrow host ranges; many are suitable only for use in species of the Solanaceae and, until very recently, there was no VIGS system suitable for use in malvaceous species such as Gossypium hirsutum. Recently a CLCrV-based vector has been produced. The usefulness of this vector for silencing in cotton was demonstrated by silencing CHL1 and PDS (Tuttle et al., 2008). However, due to the risk of the escape of this New World virus, it is inadvisable to use this vector in Pakistan where the whitefly vector is indigenous and cotton is grown widely.

Transformation of cotton is not straight forward (Wilkins, Rajasekaran, and Anderson, 2000) and a transient gene manipulation system is needed to overcome this hurdle. In this regard, the alphasatellite would seem to offer some potentially useful characters as a VIGS vector. It is small and its size could potentially be increased, possibly up to 2800nt. It has the capacity to associate with many begomoviruses and even curtoviruses, overcoming many of the host range limitations of other vectors. Here I have investigated the potential for using an alphasatellite as a VIGS vector.

4.2 Methodology

4.2.1 Modification of the Cotton leaf curl Multan alphasatellite for use as a VIGS/Expression vector

4.2.1.1 Plasmid construction

Three constructs were produced based upon the Cotton leaf curl Multan alphasatellite (CLCuMA-[PK:Fai1:98] AJ132344; Mansoor, 1999). These modifications were introduced by PCR-mediated mutation using the oligonucleotide primers detailed in Table 2.1. The primers were designed to introduce a restriction endonuclease recognition sequence, either immediately 5’ of the Rep coding sequence (yielding CLCuMA\textsuperscript{N} with an MluI site; using primers DNA1\textsuperscript{N} F and DNA1\textsuperscript{N} R) or immediately 3’ of the Rep coding sequence (yielding CLCuMA\textsuperscript{C} with an MluI site; using primers DNA1\textsuperscript{C} F and DNA1\textsuperscript{C} R). A further mutant, CLCuMA\textsuperscript{TA}, was produced with the A-rich sequence of the alphasatellite replaced with a ClaI recognition sequence (using primers DNA1\textsuperscript{mut} F & DNA1\textsuperscript{mut} R). PCR reactions used standard conditions and the recircularised (on a BamH1 restriction site) insert of the
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

alphasatellite clone as the template. Resultant PCR products were cloned into pTZ57R/T (Fermentas) and the clones were sequenced to ensure that there were no additional mutations introduced by the mutagenesis procedure.

4.2.1.2 Origin and cloning of gene/gene fragments for silencing/expression

Approximately 200-bp fragment of GFP was amplified from *N. benthamiana* line 16c (Ruiz, 1998) and inserted into the partial direct repeat construct of CLCuMA<sup>ΔA</sup> to make it CLCuMA<sup>ΔA/frGfp</sup> using primers GFPF and GFPR (Table 2.1). A similar strategy and size of ChlI gene was amplified from *G. hirsutum* (Genbank accession EU541445) to produce CLCuMA<sup>ΔA/frchl1</sup> and CLCuMA<sup>C/frchl1</sup> at *Mlu*I site using primers Chl1182F and Chl1427R. For the construction of the expression vector based on CLCuMA a 729-bp fragment from psmRSGFP (accession no. U70496; Davis SJ, 1998) was amplified and inserted into the partial direct repeat construct of CLCuMA<sup>C</sup> using primers GFP56F and GFP785R to make it CLCuMA<sup>C/GFP</sup>.

4.2.1.3 Production of constructs for *Agrobacterium*-mediated inoculation

A partial direct repeat construct of CLCuMA<sup>ΔA</sup> was prepared by digesting the clone with *Cla*I and *Bam*HI to release an approx. 1120bp fragment which was ligated into the binary vector pGreen0029. The pGreen0029 clone was then digested with *Cla*I and the full length CLCuMA<sup>ΔA</sup> insert was ligated into this as a *Cla*I fragment. A similar strategy was used to produce partial direct repeat constructs of CLCuMA<sup>N</sup> and CLCuMA<sup>C</sup> using *Mlu*I and *Bam*HI restriction sites.
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

Figure 4.1 Diagram of alphasatellite showed gene Rep (replication associated protein) is required for the replication of the component in the plants; the location of the A-rich region is also indicated. For CLCuMAΔA vector, the a-rich region was deleted with ClaI restriction site was introduced and in the second alphasatellite vector MluI restriction site was introduced with out deleting any region of the alphasatellite at the C and the N terminal of the rep. The cDNA fragments of GFP, ChlI and full length GFP were inserted to produce the CLCuMAΔA/frgfp, CLCuMAΔAfchlI, CLCuMAC/fchlI a gene silencing and CLCuMAC/gfp expression vectors.
Table 4.3 Results of the inoculation of *N. benthamiana* with CLCuMA (and its derivatives) in the presence of distinct helper viruses

<table>
<thead>
<tr>
<th>Inoculum*</th>
<th>Number of plants contained alphasatellite/ number of plants infected/ Number of plants inoculated</th>
<th>Exp.1</th>
<th>Exp.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLCuMV + CLCuMB + CLCuMA</td>
<td>6/6/6</td>
<td>5/5/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMB + CLCuMA(^{ΔA})</td>
<td>6/6/6</td>
<td>5/5/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMB + CLCuMA(^{N})</td>
<td>0/5/6</td>
<td>0/6/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMA(^{C})</td>
<td>6/5/6</td>
<td>3/4/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMA</td>
<td>3/4/6</td>
<td>5/5/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMA(^{ΔA})</td>
<td>5/5/6</td>
<td>4/5/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMA(^{N})</td>
<td>0/4/6</td>
<td>0/4/6</td>
<td></td>
</tr>
<tr>
<td>CLCuMV + CLCuMA(^{C})</td>
<td>3/5/6</td>
<td>4/4/6</td>
<td></td>
</tr>
<tr>
<td>ToLCNDV(^{*}) + CLCuMA</td>
<td>3/5/6</td>
<td>2/6/6</td>
<td></td>
</tr>
<tr>
<td>ToLCNDV(^{*}) + CLCuMA(^{ΔA})</td>
<td>4/5/6</td>
<td>4/6/6</td>
<td></td>
</tr>
<tr>
<td>ToLCNDV(^{*}) + CLCuMA(^{N})</td>
<td>0/6/6</td>
<td>0/5/6</td>
<td></td>
</tr>
<tr>
<td>ToLCNDV(^{*}) + CLCuMA(^{C})</td>
<td>3/5/6</td>
<td>2/5/6</td>
<td></td>
</tr>
<tr>
<td>CpCDPKV + CLCuMA</td>
<td>0/3/6</td>
<td>0/4/6</td>
<td></td>
</tr>
<tr>
<td>CpCDPKV + CLCuMA(^{N})</td>
<td>0/5/6</td>
<td>0/4/6</td>
<td></td>
</tr>
<tr>
<td>CpCDPKV + CLCuMA(^{ΔA})</td>
<td>0/3/6</td>
<td>0/4/6</td>
<td></td>
</tr>
<tr>
<td>CpCDPKV + CLCuMA(^{C})</td>
<td>0/4/6</td>
<td>0/5/6</td>
<td></td>
</tr>
</tbody>
</table>

* Both components of ToLCNDV were inoculated.
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

4.3 Results

4.3.1 Production and testing of constructs

To investigate the potential for using an alphasatellite as the basis for a VIGS vector, three distinct mutants were produced using the CLCuD-associated alphasatellite (CLCuMA-[PK:Fai1:98]; AJ132344) as the template. The first involved the insertion of an MluI restriction site immediately upstream of the ATG codon of the Rep gene of the alphasatellite (yielding CLCuMA\textsuperscript{N}) (Fig. 4.1). The second involved the introduction of an MluI restriction site immediate downstream of the Rep gene (CLCuMA\textsuperscript{C}) (Fig.4.1). Both these were intended to allow the insertion of foreign sequences which would be expressed as a transcription fusion. Thus the idea is to obtain the foreign sequences transcribe into the mRNA of the Rep gene.

The third construct, CLCuMA\textsuperscript{AA}, involved replacing the A-rich region with a ClaI restriction site (Fig. 4.1). There were two reasons for trying this; to see whether foreign sequences at this distance from the end of the Rep gene could be incorporated in the Rep mRNA, and thus induce silencing, and to investigate whether an alphasatellite is infectious to plants with the A-rich sequence deleted (thereby potentially increasing the size of a foreign insertion that could be tolerated, see later discussion).

The three constructs were agroinoculated to \textit{N. benthamiana} with CpCDPV, ToLCNDV and CLCuMV (in both the presence and absence of CLCuMB) (Table 4.3). For these inoculations the original CLCuMA was inoculated as a control. None of the constructs made from cotton leaf curl alphasatellite was maintained by CpCDPV and CLCuMA\textsuperscript{N} was not maintained by any of the helper viruses In light of these finding CLCuMA\textsuperscript{N} was not used for the further studies (Table 4.3).

The two remaining constructs, as well as CLCuMA, were efficiently maintained by CLCuMV, both in the presence and absence of CLCuMB, and by ToLCNDV (Table 4.3). This indicates that the introduced mutations, in each case, have not affected the ability of the molecule to replicate autonomously and be moved \textit{in trans}, by the respective viruses.
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

The symptoms induced in each case (16 to 18 days for CpCDPV, 22 to 25 days for ToLCNDV, after three weeks for CLCuMV and 14 to 16 for CLCuMV in the presence of CLCuMB) were indistinguishable from the symptoms in the absence of the alphasatellite (or mutated alphasatellite. This is in agreement with earlier reports which demonstrated that the presence of an alphasatellite does not significantly affect symptoms, although possibly a slight reduction in virus DNA levels was detected (personal observation).

4.3.2 Alphasatellite vector-mediated silencing of GFP

Initial investigation of the possibility of adapting CLCuMA for use as a silencing vector used the well established GFP silencing system that is based upon N. benthamiana line 16c (Ruiz, 1998). This contains a stable GFP transgene expressed under the control of the constitutive promoter 35S derived from Cauliflower mosaic virus (CaMV). 16c plants appear green under UV illumination due to GFP fluorescence (Fig. 4.2 panel I). In contrast, (non-transgenic) N. benthamiana fluoresces red under UV illumination due to chlorophyll autofluorescence (Fig. 4.2 panel L).

Inoculation of CLCuMAΔA/frgfp, in the presence of CLCuMV, to 16c led to typical symptoms of infection at approximately 24 days. Maintenance of the alphasatellite construct was shown by PCR amplification in 13 plants of 18 inoculated. Under UV illumination, silencing at the site of inoculation was evident at 13 days post inoculation (dpi). This appeared as a red area around the infiltration site (Fig. 4.2, panel A). Within 18 dpi, some silencing was present on leaves that were developing at the time of inoculation. This appeared as silencing of the GFP along the veins. Silencing in these leaves was not complete. The leaves above this, developing after inoculation, initially showed silencing along the veins but this spread to the whole leaf surface (Fig. 4.2, panel B). All subsequently developing tissues were entirely silenced; thus appearing red under UV illumination. The silencing was persistent, remaining until the plants senesced.

A similar result was obtained when CLCuMAΔA/frgfp was inoculated in the presence of CLCuMV and CLCuMB. In this case the symptoms of infection appeared after 14 to 16 days. Symptoms were more severe, as is typical of infections in the presence of a betasatellite (Briddon et al., 2001). In this case maintenance of the
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

Alphasatellite vector was detected in 5 plants of 6 inoculated. Silencing progressed more rapidly and was more complete than for the plants inoculated only with CLCuMV. Tissues at the top of the plant showed initial silencing at 13 days post inoculation, in comparison to 18 days for the infection in the absence of CLCuMB (Fig. 4.2, panel C&D).

In contrast, inoculation of CLCuMA$^{ΔA/fgfp}$ with ToLCNDV DNA A and DNA B yielded relatively mild infection with less pronounced leaf curling. Also the silencing induced was significantly delayed over infections involving CLCuMV. For ToLCNDV, initial symptoms of infection appeared 21 to 24 dpi. Again there was a halo of silencing visible around the site of inoculation at approx. 17 dpi. This spread to non-inoculated tissues by 20 dpi (Fig. 4.2, panel E&F). However, the plants did not fluoresce as brightly red as those inoculated with CLCuMV, suggesting that silencing was not as complete (some GFP fluorescence remaining). Maintenance of the alphasatellite vector was detected in 10 of 18 inoculated plants by PCR.

4.3.3 Alphasatellite vector-mediated silencing of an endogenous gene

As the transgene are the most favorable to gene silencing than the endogenous genes (Fagard and Vaucheret, 2000; Ruiz, 1998). So it was decided to test the ability of alphasatellite base vector CLCuMA$^{ΔA/chl}$ to silence the endogenous gene.

Inoculation of CLCuMA$^{ΔA/chl}$, in the presence of CLCuMV and CLCuMB, to N. benthamiana led to typical symptoms of infection at approximately 14 to 16 days. Maintenance of the alphasatellite construct was shown by PCR amplification in 15 plants of 18 inoculated. Silencing at the site of inoculation was evident at 10 days post inoculation (dpi). This appeared as a white area around the infiltration site (Fig. 4.2, panel G). Within 13 dpi, some silencing was present on leaves that were developing at the time of inoculation. This appeared as bleaching along the veins. Silencing in these leaves was not complete. The leaves above this, developing after inoculation, initially showed silencing along the veins but this spread to the whole leaf surface. All subsequently developing tissues were entirely silenced; thus appearing white. The silencing was persistent, remaining until the plants senesced.
In comparison, in the presence of CLCuMV and CLCuMB inoculation of CLCuMA<sup>Cfrchl1</sup>, to <i>N. benthamiana</i> a similar result was obtained. Also the silencing induced was delayed a few days over infections involving CLCuMA<sup>Cfrchl1</sup>. For CLCuMA<sup>Cfrchl1</sup>, initial silencing appeared 11 dpi. There was a halo of silencing visible around the site of inoculation at approx. 13 dpi. This spread to non-inoculated tissues by 16 dpi (Fig. 4.2, panel H). However, the plants did not show much silencing in case of CLCuMA<sup>Cfrchl1</sup> than CLCuMA<sup>Δfrchl1</sup>, suggesting that silencing was not as complete (some <i>Chl</i>1 remaining). Maintenence of the alphsatellite vector was detected in 13 of 18 inoculated plants by PCR.
Figure 4.2 Transgene GFP and endogenous gene Chl1 silencing in 16c and wild type N. benthamiana plants. (A,B) Inoculated and systemic leaf show silencing of 16c photographed in UV light inoculated with CLCuMAΔA/frgfp and CLCuMV. (C,D) Systemic silencing of 16c inoculated with CLCuMV, CLCuMB and CLCuMAΔA/frgfp (E,F) GFP silencing in association with bipartite begomovirus ToLCNDV DNA A, DNA B and CLCuMAΔA/frgfp (G,H) Use of CLCuMAΔA/Chl1 and CLCuMAΔ/Chl1 vector-based VIGS to suppress expression of endogenous gene in N. benthamiana chl1 silencing. (I) 16c GFP transgenic line. (J) N. benthamiana plant agroinoculated with only CLCuMV and CLCuMB. (K,L) Normal N. benthamiana under visible light and under UV light.
4.4 Alphasatellite mediated expression of GFP in plants

To investigate the potential for using CLCuMA as an expression vector, a full length GFP gene was inserted in sense orientation into CLCuMA\(^C\) to yield CLCuMA\(^{C/GFP}\). This construct was agroinoculated to *N. benthamiana* in the presence of CLCuMV and CLCuMB. Typical symptoms of infection CLCuMV in the presence of CLCuMB appeared approximately 14 to 16 dpi. Control plants inoculated with only CLCuMV and CLCuMB, when viewed under UV illumination, fluoresced red due to chlorophyll autofluorescence. In contrast, *N. benthamiana* inoculated with CLCuMV, CLCuMB and CLCuMA\(^{C/gfp}\) fluoresced green under UV illumination at the site of inoculation. This fluorescence was first visible approx 13 days post inoculation and at that time symptoms of infection were not visible on the plants. However, under the confocal microscope GFP fluorescence was clearly visible in tissues developing after inoculation. In a leaf which was just expanding at the time of inoculation (Fig 4.3, panel A) faint GFP fluorescence is visible. In subsequently developing leaves clear GFP fluorescence was evident (Fig 4.3, panel B). This was closely associated with the veins and did not extend very far away from vascular tissues. This is likely due to the fact that monopartite begomoviruses are phloem-limited (Rojas et al., 2001) Although this has yet to be shown for CLCuMV in the presence of the CLCuMB, the symptoms induced (enations and leaf-like outgrowths on the major veins on the undersides of leaves, Briddon et al., 2001) would suggest that this is probably the case.
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

Figure 4.3  Detection of GFP fluorescence in *N. benthamiana* inoculated with CLCuMV, CLCuMB and CLCuMA\(^{\text{C/gfp}}\). (A) green fluorescence in a leaf which was developing at the time of inoculation. (B) green fluorescence in a leaf which emerged subsequent to inoculation. Samples were taken approx. 21 days after inoculation.

4.5 Discussion

The results presented here indicate that an alphasatellite can be used as the basis for a vector for VIGS. Both CLCuMA\(^{\text{AAlfr/gfp}}\) and CLCuMA\(^{\text{Cfr/gfp}}\), carrying fragments of the *gfp* gene in antisense orientation, were capable of silencing GFP expression in *N. benthamiana* 16C when maintained by begomoviruses. Although we were unable to show either a reduction in *gfp* mRNA levels, or the production of siRNA complimentary to *gfp* (the techniques required, northern blotting and RNA hybridization, were not available at NIBGE at the time the work was conducted) the reduction in GFP fluorescence in plants infected with the CLCuMA vectors carrying fragments of *gfp*, contrasting with the bright GFP fluorescence in plants only inoculated with the helper viruses, is indicative of RNA silencing.

To further validate the vectors, a fragment of *ChlI* was introduced into CLCuMA\(^{\text{C}}\) and inoculated to *N. benthamiana*. This yielded bleaching in systemic leaves characteristic of the loss of chlorophyll, providing a visible marker for gene silencing. The results confirmed that the vector is capable of initiating silencing of an endogenous gene in *N. benthamiana.*
The reason for the inability of CLCuMA\textsuperscript{N} to systemically infect plants, in the presence of a begomovirus, is unclear. Although only minimal changes were introduced into the alphasatellite sequence (to introduce an MluI restrictions site) it is possible that this adversely affected the transcription or translation of the Rep gene. Although little is yet known about the maintenance of alphasatellites by begomoviruses, it is likely that high-level replication of these molecules is required for their maintenance, which depends upon its own Rep. There is no evidence for a (strong) selection mechanism for maintenance of alphasatellites. For betasatellites the βC1 gene is an essential factor for infection of the helper begomoviruses; it up-regulates virus titres and is a host range determinant (Saeed, 2005). There is thus a strong selection pressure (in most cases at least one case where this is not has been documented; (Li et al., 2005) for maintenance of betasatellites. Maintenance of alphasatellites thus possibly is a “numbers game”, plants containing such high levels of the satellite that it is almost always transmitted to the next cell by the virus-encoded movement proteins or to the next plant by the vector of the helper begomovirus. Thus it is likely that the changes introduced into CLCuMA\textsuperscript{N} affected Rep expression, either abolishing Rep expression, or downregulating such that there was no longer sufficient alphasatellite replication for it to be maintained during systemic infection.

The inoculations studies with CLCuMA\textsuperscript{ΔA} indicate that the A-rich sequence is not required for the infectivity or maintenance of CLCuMA. The deleted alphasatellite was maintained in plants in the presence of a begomovirus. This is in agreement with earlier studies that deleted the A-rich sequences from a betasatellite. Betasatellites have a similar A-rich sequence and this was shown not to be required for trans-replication or maintenance (Tao, 2004). The function of the A-rich sequence remains unclear. Upon their first identification it was suggested that the A-rich sequence is merely a “stuffer” required to raise the size of a nanovirus component (from which alphasatellites are believed to have evolved; Briddon and Stanley, 2006) to half (~1400bp) that of a begomovirus component (~2800). Geminiviruses have a strict size control mechanism which is believed to act during both movement in plants (mediated by the movement associated proteins; Gilbertson et al., 2003) and for encapsidation; monomeric particles containing half size molecules (Frischmuth et al.,
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

During prolonged infection of *N. benthamiana*, the CLCuMAΔA, in the presence of CLCuMV and CLCuMB, was found to increase in size, suggesting that this molecule is not stable. The precise changes that occurred to yield this size increase were not determined. It will be interesting, in the future, to investigate this phenomenon to determine what mechanism yields the size increase. This may provide information on the movement of these sub-genomic molecules in plants and their interactions with their helper begomoviruses.

One of the advantages of an alphasatellite vector, over many of the other vectors, is that it can, at least in theory, be used with any begomovirus or even curtovirus (Saunders et al., 2002). Thus it can be used to investigate gene expression (or as an expression vector) on the entire host range of the begomoviruses/curtoviruses. The experiments conducted here illustrate this possibility by showing that GFP can be silencing using both a bipartite begomovirus (ToLCNDV) and a monopartite begomovirus (CLCuMV; in both the presence or absence of a betasatellite) in *N. benthamiana*. The ability of a modified alphasatellite to induce gene silencing using several different begomoviruses has also recently been shown by (Huang, Xie and Zhou, 2009). In the context of Pakistan, cotton is the main crop species of interest and it is hoped that the vectors produced here will be useful for silencing/expression in this species. However, at this time, we have no inoculation system for this host, which precluded our testing the vectors in this plant during the study reported here.

Although the work conducted here has shown the usefulness of an alphasatellite vector for gene silencing and expression, one major hurdle remains to be overcome before their full potential can be realized. As mentioned earlier, one of the main requirements for a useful gene silencing vector is that it does not induce appreciable symptoms, which could interfere with interpretation of the silenced phenotype. All the viruses used here induce significant symptoms. For the betasatellite-associated begomoviruses, not including the betasatellite (as was shown) can reduce the symptom severity, but it does not entirely abolish symptoms. The next step is thus to identify means of either abolishing, or at least diminishing, the symptoms induced by potential helper viruses, without significantly affecting their infectivity and ability to systemically move in plants. Ongoing studies into the
Investigation of the Use of an Alphasatellite as a Silencing and Expression Vector

Pathogenicity determinants of begomoviruses may provide possible means to consistently down regulate symptoms induced by begomoviruses which would make it possible to modify potential helper begomoviruses for silencing studies using the alphasatellite vectors and the βC1 replacement betasatellite vectors produced by other researchers.

GFP has been used extensively to study the movement of viruses in plants. CP replacement vectors have, for example, been used to study the movement of BDMV (Sudarshana et al., 1998). These studies made use of the fact that bipartite begomoviruses do not require the CP for systemic infection of host plants (Stanley and Townsend, 1986; Gardiner et al., 1988; Azzam et al., 1994). However, for monopartite geminiviruses, including monopartite begomoviruses, the CP is essential for infectivity, it likely providing some essential movement functions and the implication being that movement of these viruses is either as virions or as CP-DNA complexes (Boulton et al., 1989; Briddon et al., 1989; Rigden et al., 1993). This thus precludes the CP replacement strategy for expressing GFP for virus movement studies. An alternative strategy, which may be useful for monopartite begomoviruses and curtoviruses, is to use an alphasatellite vector expressing the GFP. This would overcome any effects of mutation of the virus to facilitate GFP expression. Even for bipartite begomoviruses mutation of the CP has deleterious effects which may interfere with the interpretation of the results; for example, the time to symptom appearance is extended for bipartite begomoviruses with mutated CP (Klinkenberg et al., 1989). It is thus anticipated that the vector produced here will be useful for studying the movement of monopartite begomoviruses and curtoviruses in planta. Possibly the first use of the vector will be to study the difference between monopartite begomoviruses in the presence and absence of betasatellites, since the βC1 gene encoded by betasatellites has been shown to have possible virus movement functions (Saeed et al., 2007).
Chapter 5
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman

Foreword

As part of my Ph.D. studies I had the opportunity to spend six months in the lab of Dr. J.K. Brown (University of Arizona, Tucson, USA). The work ongoing there had cloned and sequenced begomoviruses, betasatellite and alphasatellite components from tomato samples originating from Oman (this work was conducted by Dr. A. Idris, a co-worker of Dr. Brown). Upon my arrival I was tasked to assist with the sequence analysis and to produce constructs for infectivity analysis and hence to investigate the affects of the identified alphasatellite on virus infection of plants by Agrobacterium-mediated inoculation. However, to allow a full discussion of the findings, it is necessary to provide a full account of the project, including work that was not conducted by me. The earlier work of Dr. Brown and Dr. Idris is thus duly acknowledged and I take no credit for it.

5.1 Introduction

Tomato leaf curl disease (ToLCD) and Tomato yellow leaf curl disease (TYLCD) are generic names describing diseases of tomato with symptoms such as curling (epinasty), yellowing of leaves, reduction of leaflet size, plant leaf clustering and stunting. These diseases have emerged as the main constraint to field tomato production in the tropics and subtropics and greenhouse tomato cultivation in the temperate regions. These are caused by several whitefly-transmitted begomovirus species (Czosnek and Laterrot, 1997).

These dicot-infecting viruses have a non-coding region of about 200 nt between the Rep and precoat genes called the intergenic region (IR), which contains replication and regulatory elements. Some of these monopartite begomoviruses, such as AYVV, fail to induce bona fide disease symptoms in their field hosts in the absence of their cognate betasatellite. A second group of satellite DNAs, closely related to the Rep-encoding components of nanoviruses, is associated with monopartite begomoviruses and betasatellites in the Old World. These components are now collectively referred to as alphasatellites. Previously the association of a begomovirus and betasatellite, specifically Tomato yellow leaf curls virus Oman (TYLCV-OM) and Tomato leaf curl betasatellite (ToLCB) with TYLCD in Oman has
been shown (Khan et al., 2008). These tomato diseases are considered major constraints for tomato production in Oman and are increasing in importance in the region.

5.2 Methodology

5.2.1 Virus source

Six tomato samples were collected from tomato fields in Al-Batinah region, Oman, during the 2005 tomato growing season. The tomatoes exhibited variable symptoms of leaf curling, yellowing, reduced leaflet size, and stunting. Total DNA was isolated from tomato leaf samples using the CTAB method (Doyle and Doyle, 1987). The presence of a begomovirus genome was confirmed by PCR using degenerate primers (prAV2644 and prAC1154) designed to amplify a fragment of the coat protein (Idris and Brown, 1998).

5.2.2 Cloning of begomovirus genomes and satellite DNAs

Total DNA isolated from symptomatic tomato plants was used as a template to amplify begomoviral genome and satellite DNA molecules by RCA (Inoue-Nagata et al., 2004). Amplification of circular DNA molecules was carried out using the TempliPhi 100 Amplification kit (Amersham Biosciences, Piscataway, NJ) (Khan et al., 2008; Idris, 2007). The RCA product was digested with several restriction endonucleases to identify unique sites for cloning geminivirus components (data not shown). Based on the obtained information, fragments of ~ 2.7 kb and ~1.3kb obtained with NcoI were ligated into pGEM5Zf+ vector (Promega, Madison, WI). Recombinant plasmids carrying the expected size insert were selected for sequencing. The inserts were completely sequenced using the primer walking strategy.

5.2.3 Production of constructs for Agrobacterium-mediated inoculation

Two recombinant plasmids carrying begomovirus genomes, pBatinah2.2 and pBatinah2.3, were selected for the production of constructs for Agrobacterium-mediated inoculation. The multiple fragments ligation approach was used to produce partial direct repeat constructs in the binary vector pG0029 (Hellens, Mullineaux, and Klee, 2000). pBatinah2.2 was digested with NcoI and XbaI to obtain a 2481 bp
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman

fragment, and with NcoI and XhoI to obtain a 2500 bp fragment. The binary vector pGreen0029 was digested with XbaI and XhoI to obtain a 4569 bp fragment. The three fragments were ligated and recombinant plasmids were screened by double digestion with XbaI and XhoI to release a 4981 bp fragment. The recombinant plasmid pG-Batinah2.2 was selected for biological characterization. pBatinah2.3 was digested with NcoI and BamHI to obtain a 923 bp fragment, and with NcoI and PstI to obtain a 2259 bp fragment and pG0029 was digested with BamHI and PstI to obtain a 4624 bp fragment. The three fragments were ligated and recombinant plasmids were screened with double digestion with BamHI and PstI to release a 3182 bp fragment. The recombinant plasmid pG-Batinah2.3 was selected for biological characterization.

The clone pBatinah4.3 carries a full-length alphasatellite (TYLCD-B4.3) and clone pBatinah01 carries the full-length betasatellite (TYLCβ01-Om) (Khan et al., 2008). These were selected to produce dimeric clones using the partial digestion approach (Stenger et al., 1991). The tandemly repeated copies for alphasatellite and betasatellite were released with ApaI/SacI and NcoI/SacI, respectively, individually cloned into the binary vector pGSA1403, using ApaI/SacI and NcoI/SacI. The recombinant plasmids were screened with the subcloning enzymes and pGSA-Batinah4.3 and pGSA-Batinah01 were selected for biological characterization.

5.2.4 Agro-infiltration and assessment of inoculated plants

pG-Batinah2.2 and pG-Batinah2.3 were individually introduced into Agrobacterium tumefaciens (GV3101) and pGSA-Batinah4.3 and pGSA-Batinah01 were introduced individually into the Agrobacterium strain LBA4404 and plated on medium containing antibiotic selection (Chapter 2, section 5.2.4). Empty plasmids, pGreen0029 and pGSA1403 were introduced into electro-competent A. tumefaciens cells of strains GV3101 and LBA4404, respectively, to serve as negative controls. Tomato (Solanum lycopersicum) and N. benthamiana seedlings were inoculated by slightly puncturing the leaf with a clean pin and infiltrating the bacterial suspension into the leaf stab using a disposable sterile 12cc syringe without a needle. For inoculations with more than one Agrobacterium construct, equal amounts of bacterial suspensions were mixed and infiltrated into the seedlings. The agro-infiltrated plants
were monitored for the appearance of symptoms in an insect-free and secured growth room at 28°C with a daily cycle of 14 h light and 10 h dark for four weeks.

5.3 Results

5.3.1 Sequence analysis

The complete nucleotide sequences of cloned full-length helper viruses, pBatinah2.2, pBatinah2.3 and pBatinah3.4, were determined and deposited in GenBank. The nucleotide sequences of clones pBatinah2.3 and pBatinah2.6 shared more than 94% nucleotide identity (Table 5.1) with TYLCV-OM [DQ644565] and are thus considered new variant of this virus and named TYLCV-B2.3 and TYLCV-B2.6. However, the nucleotide sequence (2763 nt) of pBatinah2.2 shared 86.5% nucleotide identity with its closest relative, TYLCV-B2.6. Nevertheless, Batinah2.2 and TYLCV-B2.6 were cloned from a single plant. Based on the presently applicable species demarcation threshold of 89%, this indicates that Batinah2.2 represents a newly identified begomovirus species, for which the name Tomato leaf curl Oman virus (ToLCOMV) is proposed (Fig. 5.3).

The genome organization of ToLCOMV is similar to other monopartite begomoviruses, encoding six conserved ORFs (V1, V2, C1, C2, C3, and C4; Table 5.1). The IR sequence of ToLCMOV is also conserved, 289 nucleotides, which contains the stem-loop structure and one direct repeat (5′-GGGGAG-3′) between nucleotide coordinates 2630 and 2634 and two direct repeats (5′-GGGGAGCTCTGGGA-3′) between nucleotide coordinates 2656 and 2668, a TATA box between nucleotide coordinates 2675 and 2679 and inverted nucleotide (5′-TCCCC-3′) between coordinates 2686 and 2689. The iterations are identical to Papaya leaf curl virus (PapLCV) Tomato leaf curl Mali virus (ToLCMLV) and Tomato leaf curl Mayotte virus (ToLCYTV). Structurally the Rep-binding arrangement for ToLCOMV is similar to other monopartite begomoviruses from the region, which consists of three direct repeats and one inverted repeat up stream of the stem-loop (Argüello-Astorga, 1994; Khan et al., 2008).
Table 5.1  Percent nucleotide sequence identities for comparisons of the full-length genome, and open reading frames of ToLCOMV with selected monopartite begomoviruses.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Full-length sequence</th>
<th>Open reading frames*</th>
<th>Predicted Rep binding motifs#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C1</td>
<td>C4</td>
</tr>
<tr>
<td>TYLCV-B2.6</td>
<td>86.1</td>
<td>77.5</td>
<td>58.8</td>
</tr>
<tr>
<td>TYLCV-OM</td>
<td>83.4</td>
<td>76.2</td>
<td>60.4</td>
</tr>
<tr>
<td>PapLCV-[PK:Lah:04]AM404179</td>
<td>74.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToLCSDVYE[YE:Tih:06]EF110890</td>
<td>75.8</td>
<td>76.1</td>
<td>64.3</td>
</tr>
<tr>
<td>TYLCV-IR[IR:Ira:98]AJ132711</td>
<td>82.1</td>
<td>72.5</td>
<td>55.7</td>
</tr>
<tr>
<td>TYLCV-Gez[SD:96]AY044138</td>
<td>81.4</td>
<td>76.7</td>
<td>62.0</td>
</tr>
<tr>
<td>TYLCV-Mld[IL:93]X76319</td>
<td>83.4</td>
<td>77.1</td>
<td>62.7</td>
</tr>
</tbody>
</table>

* Values in bold indicate the highest nucleotide sequence identity for each ORF.

# The predicted Rep binding motifs of ToLCOMV are 5’-GGGGACTCTGGGGA-3’
5.3.2 Alphasatellite

The complete nucleotide sequences of six cloned alphasatellite molecules were determined. These were found to be identical and therefore only the sequence of pBatinah4.3 was deposited in the databases. Like other alphasatellite molecules reported an association with begomoviruses, TYLCD-B4.3 is about half the size (1379 nt) of the helper begomovirus.

Analysis of the nucleotide sequence of this molecule revealed a structure similar to previously reported alphasatellites with a putative stem-loop structure with the highly conserved nonanucleotide (TAGTATTAC) sequence forming part of the loop. Search of the genome for ORFs showed the presence of a single large potential coding sequence that encodes a rolling-circle replication initiator protein (a replication-associated protein [Rep]) in the positive strand between nucleotide coordinates 75 and 941. A second ORF embedded in the Rep was also found in the same strand between nucleotide coordinates 127-612. This encodes a predicted 162aa protein and is conserved between the AYVV-associated DNA 2 (Saunders et al., 2002) and that identified from Oman. However, this has no significant similarity to the additional ORF identified in some BBTV master Rep encoding components (Beetham et al., 1997) nor yields any significant hits in Blast comparisons to the databases.

The sequence of TYLCD-B4.3 shared the highest nucleotide identity (90%) with DNA 2, an unusual alphasatellite associated with AYVV from Singapore (Saunders et al., 2002) (Fig.5.4). Based on the proposed species demarcation threshold for alphasatellites (83%; Briddon et al., 2009), the alphasatellites from Oman and Singapore are isolates of the same species. Unfortunately Briddon et al. (2009) did not consider DNA 2 identified by Saunders et al. (2002) due to its anomalous characters, and did not provide a name based upon their proposed nomenclature. Therefore we shall henceforth refer to the Oman isolate as Oman DNA 2 (OM2).

A Phylogenetic tree based upon an alignment of the full-length nucleotide sequences of the alphasatellite isolated from tomato originating from Oman and selected other alphasatellites is shown in Fig. 5.4 This shows pBatinah4.3 to segregate closely with the DNA 2 identified by Saunders et al. (2002) with very short branch lengths. However, these two sequences are very distinct from the other alphasatellites included in the analysis and were basal to them.
5.3.4 Infectivity to *N. benthamiana*

To investigate the infectivity of ToLCOMV and TYLCV-OM, partial repeat constructs of these viruses and their cognate betasatellite and alphasatellite were produced in binary vectors and introduced into *N. benthamiana* plants by *Agrobacterium*-mediated inoculation. The results of this are summarized in (Table 5.2).

*N. benthamiana* seedlings inoculated with ToLCOMV typically developed initial symptoms of infection 5-7 days post inoculation consisting of upward leaf curling, vein yellowing, reduced leaflet size, and growth stunting (Fig. 5.1, panel A). ToLCOMV in the presence of betasatellite started to show symptoms within 5-6 days post inoculation, with increased upward leaf curling, a greater reduction in leaf size and increased yellowing followed by more severe stunting (Fig. 5.1, panel B). However, in the presence of the alphasatellite, the symptoms induced by ToLCOMV were delayed, appearing at 9-10 days postinoculation. Infections in the presence of the alphasatellite had a reduced symptom severity, with less yellowing of leaves and veins compared with infections involving only the betasatellite (Fig. 5.1, panel C). On the other hand, ToLCOMV in the presence of both the satellites (alphasatellite and betasatellite), induced chlorosis following the leaf curling, yellowing of the leaflets with symptoms appearing 12-14 days postinoculation (Fig. 5.1, panel D).

*N. benthamiana* seedlings agro-infiltrated with TYLCV-OM developed initial symptoms of infection 5-7 days post inoculation. These consisted of some upward leaf curling of the leaf margins within six days post inoculation and an increased thickness of the leaves, yellowing of the leaves, stunted leaf growth and downward leaf curling (Fig. 5.1, panel E). In the presence of the betasatellite TYLCV-OM produced more pronounced symptoms and also reduced the number of days required for symptoms of infection to appear. Full symptoms appeared within 5-6 days of inoculation, consisting of downward leaf curling, deformed leaves, enhanced yellowing of leaves, reduced leaflets and the new emerging leaves appeared etiolated (Fig. 5.1, panel F). However, in the presence of the alphasatellite, TYLCV-OM symptoms had a decreased severity and symptoms took longer to appear. Leaf curling was less severe, in comparison to plants inoculated with only TYLCV-OM and the betasatellite. Symptoms appeared within 8-10 days of inoculation (Fig. 5.1, panel G). On the other hand, plants inoculated with TYLCV-OM along with the betasatellite and the alphasatellite had delayed symptoms in *N. benthamiana* plants. They started showing
symptoms between 12-14 days after inoculation. The presence of the alphasatellite decreased the symptom severity, reduced the thickness of the leaves and stunting of plants was less than in plants infected with TYLCV-OM and the betasatellite in the absence of this molecule (Fig. 5.1, panel H).

Agro-infiltration of *N. benthamiana* seedlings with ToLCOMV and TYLCV-OM induced very severe leaf curling within five to six days of inoculation. The plants progressed to show vein thickening, stunted growth and yellowing of the leaflets. (Fig. 5.1 panel I). In the presence of the betasatellite, plants inoculated with ToLCOMV and TYLCV-OM showed pronounced leaf curling, yellowing of the leaves, reduced leaflet size and early senescence (Fig. 5.1, panel J). However, in the presence of the alphasatellite, ToLCOMV and TYLCV-OM inoculated *N. benthamiana* plants exhibited reduced symptom severity. There was less yellowing of the leaves and the growth of the plants was also not as stunted (Fig. 5.1, panel K). When all four components were inoculated to plants, these developed less foliar yellowing, as well as a less pronounce reduction in leaflet size and less stunting in comparison to plants inoculated with the two viruses and the betasatellite. (Fig. 5.1, panel L).
Symptoms exhibited by *N. benthamiana* plants following inoculation with various combinations of the viruses and satellite components identified in this study. Plants were inoculated with ToLCOMV (panel A), ToLCOMV and ToLCB (B), ToLCOMV and OM2 (C), ToLCOMV, ToLCB and OM2 (D), TYLCV-OM (F), TYLCV-OM and ToLCB (G), TYLCV-OM and OM2 (H), TYLCV-OM, ToLCB and OM2 (I), ToLCOMV and TYLCV-OM (J), ToLCOMV, TYLCV-OM and ToLCB (K). ToLCOMV, TYLCV-OM and OM2 (L) and ToLCOMV, TYLCV-OM, ToLCB and OM2. The plants in panels M to O were agroinfiltrated with water (mock), *Agrobacterium* cultures containing the binary vector pGreen0029, and pGSA1403, respectively. Panel P shows a healthy *N. benthamiana* for comparison. Photographs were taken approximately after 16 days post-inoculation.
5.3.5 Infectivity to tomato

Following inoculation of tomato with ToLCOMV, initial mild symptoms appeared within 12-14 days post inoculation, consisting of very mild leaf curling (Fig. 5.2, panel A). ToLCOMV in the presence of betasatellite and alphasatellite started to show symptoms within 14-16 days of post inoculation, mild leaf curling followed the stunted systemic growth of the plants (Fig. 5.2, panel B). However, in the presence of ToLCOMV and betasatellite tomato plants initiated first symptom between 10-12 days of inoculation and the symptoms were more pronounced. After this plants were highly stunted with severe upward leaf curling and deformed leaves and reduced size of the leaves and increased yellowing (Fig. 5.2, panel C). In the presence of alphasatellite, ToLCOMV showed pronounced severe upward leaf curling, stunted plant growth and reduced leaflets within two weeks (Fig. 5.2, panel D).

Tomato seedlings agro-infiltrated with TYLCV-OM and the alphasatellite developed initial symptoms of infection within 14-16 days of inoculation consisting of some upward leaf curling, increase thickness of the leaves, yellowing of the leaves and stunted plant growth (Fig. 5.2, panel E). In the presence of both the betasatellite and alphasatellite TYLCV-OM produced very mild symptoms two weeks after inoculation, followed by upward leaf curling and yellowing of leaves (Fig. 5.2, panel F). However, in the presence of only the betasatellite TYLCV-OM started showing symptoms within 12-14 days of inoculation. Yellowing of leaves was more prominent with deformed leaves and a reduction in size of the newly emerging leaves (Fig. 5.2, panel G). The tomato plants inoculated with TYLCV-OM alone are comparatively less severe, yellowing, curling and reduction of leaflets also and it took approximately two weeks after inoculation (Fig. 5.2, panel H).

Tomato seedlings agro-infiltrated with ToLCOMV and TYLCV-OM in the presence of betasatellite gave very severe leaf curling one week after inoculation, with vein thickening, stunted growth and yellowing of the leaflets (Fig. 5.2, panel I). In the presence of alphasatellite, ToLCOMV and TYLCV-OM infections of tomato induced curling, yellowing of the leaves, reduced leaflet size and the further growth of the plants was retarded. However these symptoms were less severe than the in the presence of the betasatellite (Fig. 5.2, panel J). However, in the presence of alphasatellite (ToLCOMV, TYLCV-OM and betasatellite), tomato plants showed reduced symptom severity, yellowing of the leaves and the growth of the plants was less stunted. (Fig. 5.2, panel K). On the other hand, when both helper viruses,
ToLCOMV and TYLCV-OM, were inoculated in the absence of alphasatellite the plants showed more leaf curling and yellowing of the leaves, with more stunting. These symptoms started within two weeks of inoculation (Fig. 5.2, panel L).

Figure 5.2 Symptoms exhibited by tomato plants following inoculation with various combinations of the viruses and satellite components identified in this study. Plants were inoculated with ToLCOMV (A), ToLCOMV (B), ToLCOMV, OM2 and ToLCB (C), ToLCOMV and ToLCB (D), ToLCOMV and ToLCB (E), TYLCV-OM and MO2 (F), TYLCV-OM, OM2 and ToLCB (G), TYLCV-OM and ToLCB (H), TYLCV-OM (I), ToLCOMV, TYLCV-OM and ToLCB (J), ToLCOMV, TYLCV-OM and OM2 (K), TYLCV-OM, TYLCV-OM, OM2 and ToLCB (L), ToLCOMV and TYLCV-OM (M). The plant in panel N was infiltrated with water to act as a mock inoculated control. Photographs were taken 24 days post-inoculation.
Table 5.2  Results of the Agrobacterium-mediated inoculation of *N. benthamiana* and tomato with various combinations of the viruses and satellite components identified in this study

| Inocula* | RCA Amplification | *N. benthamiana* | Tomato | |
|----------|-------------------|------------------|--------|----------|-------------------|-------------------|----------|
|          |                   | no. of plants positive/no. plants inoculated# |         | |
|          |                   | TO | TY | B | I | TO | TY | β | I |
| TO       | +                 | 12/12 | - | - | - | 12/12 | - | - | - |
| TO+ 1    | +                 | 12/12 | - | - | 11/12 | 12/12 | - | - | 10/12 |
| TO+ β    | +                 | 12/12 | - | 12/12 | - | 12/12 | - | 12/12 | - |
| TO+ 1+ β | +                 | 12/12 | - | 11/12 | 12/12 | 12/12 | - | 12/12 | 11/12 |
| TO+ TY   | +                 | 12/12 | 12/12 | - | - | 11/12 | 12/12 | - | - |
| TO+ TY+ 1| +                 | 12/12 | 12/12 | - | 9/12 | 12/12 | 12/12 | - | 10/12 |
| TO+ TY+ β| +                 | 12/12 | 12/12 | - | - | 12/12 | 10/12 | - | - |
| TO+ TY+ 1+ β | +         | 11/12 | 11/12 | 12/12 | 11/12 | 12/12 | 12/12 | 12/12 | 10/12 |
| TY       | +                 | - | 12/12 | - | - | - | 12/12 | - | - |
| TY+ 1    | +                 | - | 12/12 | - | 12/12 | - | 12/12 | - | 11/12 |
| TY+ β    | +                 | - | 12/12 | 12/12 | - | - | 12/12 | 12/12 | - |
| TY+ 1+ β | +                 | - | 12/12 | 12/12 | 11/12 | - | 12/12 | 12/12 | 11/12 |
| Mock-inoculated control | - | - | - | - | - | - | - | - | - |
| pGSA-1403 control | - | - | - | - | - | - | - | - | - |
| pG0029 control   | - | - | - | - | - | - | - | - | - |

*Plants were examined for the presence of each inoculated component by diagnostic PCR on RCA product and each experiment was repeated three times.*

*The inocula were constructs for the infectivity of ToLCOMV (TO), TYLCV-OM (TY), ToLCB (β) and Oman DNA2 (1)*
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman

Figure 5.3  Phyllogenetic dendrogram based upon an alignment of the complete genomes of selected begomoviruses from the databases. Numbers at the nodes are percentage bootstrap confidence values (1000 replicates). Vertical lines are arbitrary and horizontal lines are proportional to calculated mutation distances. The tree is rooted on the sequence of CLCuRV-[PK:Fai1:06] AM490309. The species used were Croton yellow vein mosaic virus (CYVMV), Papaya leaf curl virus (PaLCuV), Pepper leaf curl virus (PepLCuV), South African cassava mosaic virus (SACMV), Tomato leaf curl Sinaloa virus (ToLCuV), Tomato leaf curl Uganda virus (ToLCuV), Tomato leaf curl mottle virus (ToLCMoV), Tomato leaf curl Mayotte virus (ToLCuTV), Tomato leaf curl Comoros virus (ToLCKMV), Tomato leaf curl Madagascar virus (ToLCMGV), Tomato yellow leaf curl Malaga virus (TYLCMaV) and Tomato yellow leaf curl virus (TYLCV). The isolate descriptors are as given in (Fauquet et al., 2008).

Table 5.3  The highest and the lowest percentage nucleotide sequence identities for pair wise comparisons of the sequences of the viruses from Oman with selected begomoviruses from the databases

<table>
<thead>
<tr>
<th>TYLCV-OM</th>
<th>ToLCuV</th>
<th>TYLCV-IL</th>
<th>PaLCuV</th>
<th>PepLCuV</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)*</td>
<td>(1)*</td>
<td>(12)*</td>
<td>(2)*</td>
<td>(4)*</td>
</tr>
<tr>
<td>100</td>
<td>86.66-63.2</td>
<td>98.9-83.2</td>
<td>64.5-63.2</td>
<td>74.7-74.0</td>
</tr>
<tr>
<td>100</td>
<td>92.0-82.1</td>
<td>74.8-74.6</td>
<td>73.5-72.7</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>65.3-64.4</td>
<td>75.5-73.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>89.1-63.1</td>
<td>97.1-64.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>89.1-63.1</td>
<td>74.8-74.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>74.7-74.0</td>
<td>73.5-72.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The number of sequences available which were used in the comparisons.
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman

Figure 5.4 Phylogenetic dendrogram based upon an alignment of the complete nucleotide sequences of selected alphasatellites available in the databases. The tree is rooted on the ToLCB-[OM:Bat1:08]DQ644566. The species used were Ageratum yellow vein India alphasatellite (AYVIA), Ageratum yellow vein alphasatellite (AYVA), Cotton leaf curl Multan alphasatellite (CLCuMA), Hibiscus leaf curl alphasatellite (HLCuA), Malvestrum yellow mosaic Hainan alphasatellite (MalYMHnA), Tobacco curly shoot alphasatellite (TbCSA), Tomato leaf curl China Yunan alphasatellite (TYLCYnA). The alphasatellite isolate descriptors are as given in (Briddon at al., 2009) and the sequence produced as part of this study are highlighted.
Table 5.4  The highest and the lowest percentage nucleotide sequence identities for pair wise comparisons of the alphasatellite isolated from tomato from Oman with the available sequences of other alphasatellites.

<table>
<thead>
<tr>
<th></th>
<th>DNA 2 (OM2) (1)*</th>
<th>AYVA (3)*</th>
<th>HLCA (4)*</th>
<th>OLCA (1)*</th>
<th>AYVV DNA 2 (1)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>84.9-48.4</td>
<td>32.1</td>
<td>91.9</td>
<td>DNA 2 (OM2)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>86.7-28.2</td>
<td>65.9-28.9</td>
<td>88.9-28.2</td>
<td>AYVA</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>84.7-61.7</td>
<td>33.3-30.12</td>
<td>HLCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>28.9</td>
<td>OLCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>AYVV DNA 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The number of sequences available which were used in the comparisons.
5.4 Discussion

The study presented here has shown evidence indicating that TYLCD in the Sultanate of Oman is caused by a complex of begomovirus genomic components including several variants of TYLCV-OM. In addition to these TYLCV variants another begomovirus was cloned and sequenced that shared less than 89% nucleotide identity with its closest relative (TYLCV-B2.6, a TYLCV-OM variant). Following the ICTV guidelines (Fauquet et al., 2008) the new begomovirus species was tentatively named ToLCOMV. Analysis of the full-length nucleotide sequence of ToLCOMV showed that this begomovirus species has a genomic arrangement, notably the presence of V2, similar to other begomoviruses from the Eastern Hemisphere. Although, ToLCOMV is a distinct new begomovirus species, its V1, V2, C2 and C3 ORFs shared more than 94.5% nucleotide identity with TYLCV-B2.3 which was cloned from the same plant. However, the nucleotide sequence for the C4 gene of ToLCOMV shared 94.1% nucleotide sequence identity with PapLCV (Table 5.1) from Pakistan. Moreover, the sequences and arrangement of ToLCOMV Rep-binding motifs are also similar to PapLCV. These nucleotide sequence comparison results and the recombination analysis confirmed that ToLCOMV evolved as a result of interspecific recombination between TYLCV-OM and PapLCV. The high levels of nucleotide identity between ToLCOMV and TYLCV-B2.6 suggested that this recombination event took place recently.

Apparently, this new begomovirus is not widespread, which was detected by PCR and cloned using RCA technology from one tomato sample, compared to the TYLCV-OM, which was detected and cloned from all six tomato samples. This confirmed the earlier findings (Khan et al., 2008) that the well adapted TYLCV-OM is widespread and causes most of the TYLCD in tomato. The analysis showed that the complete genome of TYLCV-OM variants had nucleotide divergence of 0.01-6.3% that constitutes a base for evolving into new strains and probably new distinct species. Inspection of the alignment of complete sequence of these TYLCV variants revealed that the nucleotide divergence was distributed throughout the genome and not restricted to the IR (data not shown). This cannot be attributed to point mutation alone and intra specific recombination might have also contributed to the divergence TYLCV variants given the fact that the most divergent variants were cloned from a single plant.

For this novel recombination to occur a mixed infection between putative progenitors, TYLCV-OM and PapLCV, must have happened in a common host. The high nucleotide
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman

identity of recombinant between the ToLCOMV and putative progenitor TYLCV-OM suggested that the recombination event took place relatively recently. Because of the arid nature of southern Arabia, it is possible that perennial weeds or ornamental plants might have served that role. PapLCV, the likely source of some of the sequences of ToLCOMV, is a begomovirus that occurs throughout the Indian subcontinent (Saxena et al., 1998; Mansoor et al., 2003). Because of the proximity and the trade between the two regions, this virus or closely related species, might have accidently been introduced in southern Arabia. The putative donor might have adapted to the new environment in southern Arabia and established in weeds or perennial ornamentals.

Although it was first reported from papaya (Saxena et al., 1998), PapLCV has also been identified in cotton with cotton leaf curl disease-like symptoms, as well as being shown experimentally to be able to cause CLCuD symptoms in cotton in the presence of the CLCuD-associated betasatellite (Mansoor et al., 2003). Some of the components of the CLCuD complex, specifically CLCuMV and the CLCuD-associated betasatellite (Briddon et al., 2003) have been identified in Hibiscus rosa-sinensis. Pakistan exports ornamental plants to the Gulf States and the source H. rosa-sinensis plants from which cuttings are taken for export are almost universally symptomatic for CLCuD-like symptoms (R.W. Briddon, personal communication). This provides a possible route for the introduction of PapLCV sequences into Oman, although it is unclear whether recombination occurred before or after introduction. The close similarity between TYLCV-OM and TYLCV-Iran (Khan et al., 2008) provides further evidence for the movement of viruses between the Middle East/southern Asia and southern Arabia.

A whitefly transmitted isolate of TYLCV-OM and its’ associated betasatellite were identified from Tomato leaf curl Oman complex particularly form the Al-Batinah region of Oman (Khan et al., 2008) from where other isolates were also found. The features of the TYLCV-OM genome, the absence of a detectable DNA-B component, and the presence of a betasatellite, suggested that the causal agent of TYLCD in Oman is a monopartite begomovirus-satellite complex (Khan et al., 2008). The genome organization of TYLCV-OM resembles to that of other Middle Eastern monopartite begomoviruses, including TYLCV from Iran (Hajimorad et al., 1996) and TYLCV from Israel (Navot et al., 1991), whose genomes differ from two other tomato-infecting begomoviruses from Asia, TYLCTHV from Thailand and ToLCNDV from India (Padidam et al., 1995), which have a bipartite
Effects of an Unusual Alphasatellite on Symptoms Induced by Begomoviruses Originating from Oman

organization. These distinct tomato-infecting begomovirus derivates from Asia and Middle East, as they have been identified in a ‘recombination or bridge zone’ situated between the Indian subcontinent and the Mediterranean- North Africa region.

The inoculation experiments indicate that ToLCOMV and TYLCV-OM act synergistically to induce severe disease. Synergism is well documented for many RNA viruses, for example many synergistic interactions between plant infecting viruses involved a potyvirus as one of the components. The basis for this has been shown to be the HC-Pro encoded by the potyvirus which acts to enhance the titer of the second virus, likely due to its encoding a strong suppressor of post-transcriptional silencing activity (Pruss et al., 1997; Kasschau and Carrington, 1998). Synergism is also well documented for begomoviruses. For example, an epidemic of a severe form of cassava mosaic virus across Africa has been attributed to a synergistic interaction between two begomovirus species, African cassava mosaic virus and East African cassava mosaic virus (Harrison et al., 1997). The synergism was later shown to be based upon these viruses encoding complementing suppressor of silencing (Vanitharani et al., 2004). Elucidating the basis for the synergism in the case here will be an interesting future area of work.

It is unusual for two begomoviruses to interact with a single betasatellite in planta. A similar situation exists with CLCuD on the sub-Continent where several begomoviruses have been shown to interact with a single betasatellite, frequently in multiple infections, to induce the disease (Mansoor et al., 2003). The selective advantage of this is unclear, although it possibly represents an intermediate situation following the introduction of a new virus species into an area. Certainly the situation on the sub-Continent has not proven stable and CLCuD is at this time caused by a single begomovirus species which is a recombinant derived from two of the earlier viruses (L. Amrao, manuscript in preparation). Possibly a similar situation will unfold in Oman with one virus gaining the upper hand. Nevertheless, as has been shown for a number of other plant diseases, the ToLCB acts to enhance disease symptoms since betasatellites encode a strong pathogenicity determinant (Saunders et al., 2004; Cui et al., 2004; Saeed et al., 2005).

Surprising is the finding that the alphasatellite associated with tomato disease in Oman ameliorates symptoms. Although earlier reports have shown alphasatellites to reduce virus and betasatellite DNA levels in plants (Mansoor et al., 1999; Saunders and Stanley
Effects of an Unusual Alphasatellite on Symptoms Induced by
Begomoviruses Originating from Oman

1999), symptom amelioration was not a feature of the interaction. At this time it is assumed that alphasatellites are merely passengers of begomovirus-betasatellite infections – their effects being mediated by interference in competing for cellular resources. The results presented here suggest that there may be more to the “story”, which warrants more detailed investigation. The alphasatellite associated with tomato disease in Oman is distinct from all other satellites that have been examined in detail. Based on the proposals of Briddon et al. (2009) it should possibly be included in a separate genus with the DNA 2 associated with AYVV originating from Singapore. These alphasatellites, unlike all others for which sequences are available, encode a second ORF overlapping the Rep which could be expressed. Possibly the product of this ORF encodes the factor responsible for symptom amelioration. The product predicted to be encoded by the additional ORF has no similarity to the product of the additional ORF of some BBTV master Rep encoding components and has no significant similarity to any sequences in the database, providing us with no clues to its possible function. Future studies should investigate the possibility that this additional ORF is functional by site directed mutation. A detailed knowledge of the mechanism of symptom amelioration could provide a novel strategy for overcoming losses due to begomoviruses in plants.

The presence of two closely related alphasatellites in Singapore and Oman is surprising. This demonstrates that we are still a long way from determining the full extent of genetic diversity of this class of molecules. The symptom amelioration phenomenon demonstrated for the alphasatellite from Oman demonstrates that, despite the earlier assumption that these components are not of significance and therefore not of interest, they have important biological functions and studying them can further our knowledge and provide possibly useful biotechnological tools.
Chapter 6
Although identified before betasatellites, the alphasatellites have not attracted as much attention from researchers. At the end 2008 there were over 260 full-length betasatellite sequences in the databases (Briddon et al., 2008), but only 68 full-length begomovirus-associated alphasatellites (the term alphasatellite also encompasses the closely related Rep encoding satellite-like molecules of nanoviruses, 23 sequences of which were available in the databases; Briddon et al., 2009). This discrepancy is understandable since the alphasatellites appeared to be, based on the available evidence, mere passengers; the helper begomovirus and, particularly, the betasatellite components of the complexes being the factors important for symptoms, host range and insect transmission (Briddon and Stanley, 2006). However, as well as being interesting with respect to their evolutionary origin, investigation of which may provide information on the mechanism and interactions that have been involved in the evolution of the begomovirus-betasatellite complexes (the precise origins of the betasatellites remains unknown), the work conducted here has shown that alphasatellites can have a significant effect on the symptoms of their associated begomoviruses and can be adapted as useful biotechnological tools.

Significant advances have been made in understanding the interactions between betasatellites and their helper viruses since betasatellites were first identified in 2000 (Saunders et al., 2000; Briddon et al, 2001). Despite being identified before betasatellites (Mansoor et al., 1999) advances in elucidating the function(s) of the alphasatellites has not kept pace, even though they, at first glance, appear relatively simple. First identified in association with the CLCuD begomovirus-betasatellite complex, the study of this disease has been most intensive with respect to the alphasatellite. The alphasatellite was found in all isolates of the disease originating from the 1990s for which the presence of this component was investigated. The work reported here (Chapter 3) has shown that this is no longer the case, with isolates from the Punjab having no associated alphasatellite and those from Sindh having an occasional associating with this group of molecules. The reason for this remains unclear. It would seem likely that the changes, in both the virus/betasatellite and the nature of the cotton varieties being cultivated (resistant and non-resistant, respectively) are responsible for these differences and for the apparent recent change in the situation in the Punjab.
The precise benefits that begomovirus-betasatellite complexes gain from the presence of alphasatellites remains unclear and this mystery is now compounded by the demonstration here that the “Burewala strain” of CLCuD does not associate with this class of molecules. The earlier “Multan strain” was invariably associated with an alphasatellite. The invariant presence of this class of satellites strongly suggests that it plays an important role, even though that role is not essential, as has been shown experimentally (Mansoor et al, 1999; Saunders et al., 1999). This likely indicates that the effects of the presence of the alphasatellite are very subtle, providing only a very slight selective advantage in the field. This is consistent with the suggestion that alphasatellites may act similarly to defective interfering DNAs, slowing down virus replication by competing for cellular resources. We might thus speculate that for the Burewala strain (and other begomovirus complexes shown here and elsewhere; Briddon et al., 2004) the alphasatellite component of the complex is no longer required. This in turn suggests that either some other component has taken over its role, or that the retarding effects of the alphasatellite have become detrimental; the virus now needing its full pathogenicity/replication potential to infect host plants. Certainly for the CLCuD complex this hypothesis is attractive since the loss of association with an alphasatellite (the appearance of the Burewala strain) coincided with resistance breaking. The need to overcome host plant resistance may thus be the underlying reason for the loss of the alphasatellite, since this is likely to have been a major hurdle for the virus. Certainly there were other changes in the CLCuD complex including the appearance of a new species of begomovirus helper (a recombinant lacking a TrAP gene) and a recombinant betasatellite. The precise effects (and need for these changes to break resistance) of these changes in the CLCuD begomovirus-betasatellite complex in response to the introduction of resistant cotton varieties are under investigation.

The second possibility, that another factor/molecule has taken over the role of the alphasatellite, is less plausible. diDNA is a common feature of most, if not all, geminivirus infections. It is thus possible that a defective molecule, derived from one of the components could be present. However, there is no indication that any one such molecule is prevalent, or in greater abundance than in the earlier Multan strain infected cotton plants (M.N. Tahir and S. Akhtar, unpublished results). It is clear that, although we are now aware of the absence of the alphasatellite, further studies are
required to define the underlying reasons for this major change in the CLCuD complex. Such studies are now in process.

The apparent lack of interest in alphasatellites means that we are far from having established the full diversity of alphasatellites. Previously unidentified alphasatellites were identified in the study of both weeds and cotton. Although all the alphasatellites isolated from Pakistan in this study (Chapter 3) are typical of the majority those identified earlier, that identified from Oman is unusual (Chapter 5). Based on the proposals of Briddon et al. (2009), it is likely that the Oman and Singapore DNA 2 alphasatellites will ultimately be classified in a separate genus from the remainder of the so far identified begomovirus-associated alphasatellites.

A more detailed investigation, using infectious clones, is needed to establish the basis for the changes that are happening in recent discovery. However, before this can happen, there is a need for an efficient infectivity system for begomoviruses of cotton. Without such a system it will not be possible to investigate the infectivity of the different virus, betasatellite and alphasatellite clones to resistant and non-resistant cotton cultivars to determine what their interaction with and to investigate the basis for resistance breaking (and whether the alphasatellite (or rather the absence of the alphasatellite) has a part to play in this. Efforts are underway to establish an infectivity system for cotton.

VIGS is a powerful technique for down-regulating the expression of gene in plants. It is now widely used as a means of investigating gene function and has been adapted for high-throughput. For example, Senthil-Kumar and Udayakumar (2006) used a Tobacco rattle virus-based VIGS vector to investigate the function of moisture stress induced genes. Such an analysis would have been far more difficult without the availability of VIGS, requiring time consuming plant transformation. The work conducted in Chapter 4 has shown that an alphasatellite can be adapted for use as a VIGS vector and additionally can be used with a variety of helper viruses, thus likely extending the plant host range of the vector. This is a very promising advance since, at least in theory, it opens up the whole host range of begomoviruses and curtoviruses to VIGS studies using a single vector. In the context of Pakistan, our main interest is cotton. Recently, Tuttle et al., (2008) demonstrated a VIGS vector based upon the bipartite begomovirus Cotton leaf crumple virus (CLCrV) vector. This is the first
vector reported that is suitable for use in the plants of family *Malvaceae* but for biosecurity reasons it is inadvisable to use this in Pakistan at this time; suitable high containment facilities are not available at this time and the country has enough problems with geminiviruses without introducing exotic viruses. Very similar results have recently been published by Huang, Xie and Zhou (2009) using the Tomato yellow leaf curl China alphasatellite. The ultimate aim of work conducted here is to use the alphasatellite for VIGS studies in cotton, the major source of foreign exchange for the economy of Pakistan. Such a vector is needed for many reasons, including investigation into biotic and abiotic stress tolerance and improvement of fibre quality.

As proof of concept this GFP was used as the marker gene. GFP has been extensively used as a marker to study the movement of viruses through plant tissues. The alphasatellite vector expressing GFP thus promises to extend such GFP marker studies to monopartite viruses (begomovirus and curtovirus) which, due to their absolute requirement for the CP (precluding the CP replacement approach to GFP expression) and size limitation (precluding the insertion of the GFP gene without deletion) for encapsidation and movement, have not been studied in such detail as the bipartite begomoviruses. This will allow a far easier analysis of gene mutations that affect virus movement to be studied and should allow much more rapid progress.

Full realization of the potential of the alphasatellite vector developed as part of the present study will have to await an efficient inoculation system for introducing betasatellite requiring begomoviruses into cotton. In addition, it is desirable for a VIGS vector to induce few if any symptoms in plants, since these could mask the silencing phenotype. For betasatellite requiring begomoviruses, such as those infecting cotton that cause CLCuD, the only suitable helper begomovirus-betasatellite combinations induce severe symptoms. The major symptom determinant of these complexes is the βC1 protein encoded by the betasatellite (Saunders et al., 2004; Cui et al., 2005; Saeed et al., 2005; Qazi et al., 2007) which, for a number of these complexes including cotton, is essential for efficient infection (Saunders et al., 2000; Briddon et al, 2001). Efforts are underway at this time to better understand the function and mechanism of action of βC1 (Cui et al., 2005; Saeed et al., 2007, 2008) which may provide a means of engineering a βC1 protein that does not induce such severe symptoms but is nevertheless able to provide the necessary functions to the helper virus. Possibly the unusual classs of alphasatellites represented by AYVV-
associated DNA 2 and the alphasatellite identified from Oman may be useful. If this symptom ameliorating alphasatellite can be adapted as a vector it may overcome all the problems discussed above. However, first it is necessary to investigate molecular basis for this phenomenon. Possibly the additional ORF identified in these molecules is responsible and efforts are now underway to answer this question.

In view of the unusual nature of Oman alphasatellite (DNA 2) (encoding a distinct Rep) and behavior (symptom amelioration) this satellite-like molecule needs further investigation. The molecular basis for symptom amelioration and the possibility that the additional ORF (overlapping the Rep genes) has a part to play in this warrants immediate investigation. Additionally the origins of these DNA 2 needs to be established. It is unusual that two such similar molecules should be identified from such geographically distinct locations. The possibility that the molecule from Oman has its origins in south East Asia (or vice versa) needs to be looked at. Both the Gulf States and Singapore are central hubs for air travel and it is thus possible that the molecule has been introduced in one (or both) areas by movement of agricultural products or accidental movement of insect vectors should be established.

Chapter 4 showed that, in addition to being useful as a VIGS vector, alphasatellites will also prove useful as expression vectors. Expression of GFP from the vector will be useful to non-destructively investigate virus movement in planta. CP replacement vectors expressing GFP have proven very useful in investigating the movement of bipartite begomoviruses (Sudarshana et al., 1998). However, the essential nature of the CP of monopartite begomoviruses has precluded a similar strategy for the investigation of the movement of these viruses. Use of a GFP expressing alphasatellite vector overcomes this limitation without modification of the helper begomoviruses and will allow a wide range of such viruses to be investigated; a “one size fits all” situation. In the first instance this vector should be used to investigate the affects the betasatellite has on begomoviruses by comparing the distribution of begomoviruses in planta in the presence and absence of the betasatellite. This will be of interest in light of the recent demonstration of possible movement functions of the βC1 gene (Saeed et al., 2007).

For the alphasatellite vector only betasatellite-associated begomoviruses are available for use as helper viruses at this time. However, the recent identification of
an unusual, *Malvaceae*-infecting bipartite begomoviruses in Vietnam and India (Ha et al., 2006, 2008; Ghosh, 2008) holds out the possibility of using these as helper viruses for the alphasatellite vector. This approach needs investigating. In the first instance, the GFP-expressing alphasatellite will be used to study the effects of begomovirus association with betasatellites (thus to examine a begomovirus infection in the presence and absence of a betasatellite) to determine whether the βC1 gene encoded by betasatellites contributes to virus movement or not; as has recently been suggested (Saeed et al., 2007).

Finally, the finding that the alphasatellite originating from Oman (Chapter 5) can ameliorate symptoms provides the possibility that this may be used to overcome one of the present problems with the alphasatellite vector; namely that all presently available helper viruses produce quite severe symptoms in plants which could mask any phenotype due to silencing. Thus a vector based upon the Oman alphasatellite could possibly act as both a vector and concomitantly ameliorate the symptoms of whatever helper virus is used to support it. This possibility needs investigation.
Chapter 7


Azzam, O., Frazer, J., de la Rosa, D., Beaver, J. S., Ahlquist, P., and Maxwell, D. P.
(1994). Whitefly transmission and efficient ssDNA accumulation of bean

curly top virus: a newly described *Curtovirus* species from southwest Texas


Baulcombe, D. C. (1996). Mechanisms of pathogen-derived resistance to viruses in


discovery mechanisms and exploitation. *Philosophical Transactions of the

resistance against virus infection. *Annual Review of Phytopathology* **28**, 451-
472.


Bedford, I. D., Briddon, R. W., Brown, J. K., Rosell, R. C., and Markham, P. G.
(1994). Geminivirus transmission and biological characterisation of *Bemisia
tabaci* (Gennadius) biotypes from different geographic regions. *Annals of
References


References


References


References


References


References


References


References


References


References


References


References


References


References


References


References


References


Publications
Tomato leaf curl disease (ToLCD) is a serious problem throughout the warmer parts of the world [7]. On the Indian subcontinent, the disease is caused by a diverse range of single-stranded DNA viruses of the genus Begomovirus (family Geminiviridae) [17]. The virus first identified causing the disease on the sub-continent in the 1990s, tomato leaf curl New Delhi virus (ToLCNDV), is a typical bipartite begomovirus; having two genomic components, designated DNA A and DNA B [15]. In recent years, a diverse range of monopartite begomoviruses (viruses which lack the DNA B component) have been shown to cause ToLCD. The majority of these viruses are associated with DNA β, a recently identified group of symptom-modulating, single-stranded DNA satellites that occur only in the Old World [5, 6].

In 2005, three tomato (Solanum lycopersicum) plants were collected in a kitchen garden in Faisalabad, Punjab Province, Pakistan. These plants exhibited an unusually severe “leaf curl” phenotype consisting of highly curled leaves, reduced in size with frond-like enations on the veins on their undersides. Total nucleic acids were extracted as described previously [8]. Initial analysis by PCR with universal primers [4] and limited sequencing showed two of the plants to be infected with ToLCNDV (results not shown). Two primer pairs, BF (5’-ACGCGT GCCGTGCTGCTGCCCCCATT GTCC-3’) = BR (5’-ACGCGT ATGGGCTGYCGA AGTTSAGACG-3’) and β01/β02 [2], were used to PCR-amplify the full-length begomovirus component and DNA β satellite, respectively, from the third plant. Multiple clones were obtained, and one clone from each amplification (designated clones SA1 and Sβ2 for the begomovirus and DNA β, respectively), was selected and sequenced in its entirety in both orientations. Sequences were assembled and analyzed with the aid of the Lasergene package of sequence analysis software (DNASTar Inc., Madison, WI, USA), and multiple sequence alignments were performed using Clustal X [18]. Phylogenetic trees were constructed using the Neighbour Joining algorithm of Clustal X and displayed, manipulated and printed using Treeview.
Sequences of begomoviruses and associated satellites used in the analyses were obtained from the sequence databases.

The complete nucleotide sequences of clones SA1 and Sβ2 (available in the databases under accession numbers AM501481 and AM490309, respectively) were determined to be 2753 and 1370 nucleotides, respectively. Sequence comparisons showed the begomovirus (SA1) to be nearly identical to all isolates of cotton leaf curl Rajasthan virus (CLCuRV) available in the databases (over 99% nucleotide sequence identity). However, the sequence of SA1 also shows over 89% identity to a number of cotton leaf curl Multan virus (CLCuMV) isolates (between 86 and 94% identity for the isolates compared here). For geminiviruses, the threshold cut-off value for distinguishing species from strains currently rests at 89% [9]. This indicates that there is some overlap between isolates of CLCuRV and CLCuMV, a situation that will need to be resolved in the future. Phylogenetic comparisons show SA1 to cluster with the CLCuRV isolates, being most closely related to an isolate originating from India (CLCuRV-[India: Abohar: 2003], AY795606). This indicates that the begomovirus identified causing severe ToLCD is an isolate of CLCuRV, for which we propose the descriptor CLCuRV-[Pakistan: Faisalabad: tomato: 2005]. Significantly, Abohar in India is only a short distance, across the border, from Faisalabad, and this suggests that CLCuRV may have spread into Pakistan from India.

CLCuRV is one of at least 7 begomoviruses shown to be associated with cotton-leaf curl disease (CLCuD) [11, 13]. CLCuD is endemic to the majority of cotton cultivating areas of central/southern Pakistan and western India. The disease was epidemic during the 1990s but was brought under control by the use of resistant cotton varieties. However, in 2001, a resistance-breaking strain of the virus appeared in central Pakistan [12] and has spread throughout most of the country since then, as well as into India. All previously characterised isolates of CLCuRV were identified from India and infected cotton. This is the first identification of CLCuRV from Pakistan and is the first time it has been identified infecting tomato.

The sequence of Sβ2 shows the typical arrangement of DNA β satellites, with a single open reading frame in the complementary sense (known as βC1; coordinates 508–155), a region of sequence rich in adenine (coordinates 719–1013), and a sequence motif highly conserved between all DNA β satellites, known as the satellite conserved region (SCR; coordinates 1265–14). The βC1 gene is predicted to encode a 118-amino-acid protein which shows greater than 98% amino acid sequence identity to the βC1 products of other DNA β satellites isolated from cotton affected by CLCuD. The satellite does not contain the recombinant SCR recently identified for the CLCuD DNA β associated with resistance breaking in cotton [1]. Phylogenetic analysis shows this DNA β to be most closely related to an isolate identified in Hisar (Haryana State, India; AY763123; 99.5% nucleotide sequence identity), which is situated close to the border with Pakistan (Fig. 1). Sβ2 is thus a typical CLCuD DNA β of the type first identified in association with the disease epidemic that occurred in the 1980s to 1990s [3].

CLCuD sporadically affects, in addition to cotton, a number of crop species including radish [14], papaya [13] and chilli peppers [10]. In most cases, the viruses causing the disease have only been poorly characterised, although in each case the presence of the CLCuD DNA β was shown. These infections of non-malvaceous hosts likely occur due to a high inoculum pressure. Faisalabad is situated in the centre of the cotton-growing area of the Punjab and, during the summer season, few other crops are in the field. It is likely therefore, that during this period, the whitefly vector, carrying the components (begomovirus and DNA β) causing CLCuD, are in abundance. This report of CLCuRV occurring in Pakistan brings to five the number of begomoviruses associated with CLCuD that have been reported in the country (these being cotton leaf curl Multan virus, cotton leaf curl Alabad virus, cotton leaf curl Kokhran virus and papaya leaf curl virus) out of a total of 7 begomoviruses identified in association with CLCuD on the subcontinent. Additionally, it shows for the first time that CLCuRV and CLCuD DNA β can infect tomato.
Fig. 1. Phylogenetic dendrograms based upon selected complete sequences of the genomes (or DNA A genomic components; left) and DNA β satellites (right). Vertical branches are arbitrary, horizontal branches are proportional to calculated mutation distances. Values at nodes indicate percentage bootstrap values (1000 replicates). Begomovirus sequences used for comparison were cotton leaf curl Alabad virus (CLCuAV), cotton leaf curl Bangalore virus (CLCuBV), cotton leaf curl Kokhran virus (CLCuKV), cotton leaf curl Multan virus (CLCuMV), cotton leaf curl Rajasthan virus (CLCuRV), okra yellow vein mosaic virus (OYVMV), tomato leaf curl Bangladesh virus (ToLCBDV), tomato leaf curl Bangalore virus (ToLCBV), tomato leaf curl Gujrat virus (ToLCGV), tomato leaf curl New Delhi virus (ToLCNDV), and tomato leaf curl Sri Lanka virus (ToLCSLV). This tree was arbitrarily rooted on the sequence of tomato mottle virus (ToMoV), a distantly related begomovirus. The DNA β satellite sequences used for comparison (indicated as the virus, if known, or alternatively the disease with which they were associated) were cotton leaf curl disease (CLCuD), Ageratum yellow leaf curl disease (AYLCuD), tomato leaf curl disease (ToLCD), and cotton leaf curl virus (TYLCCNV). The DNA β satellite was arbitrarily rooted on the sequence of CLCuD DNA 1, an unrelated sequence of similar size. The database accession numbers are indicated, and the sequences associated with tomato from Faisalabad are highlighted by black boxes. DNA β components of the “Malvaceae type” are labelled (Mal). These DNA β molecules cause disease in malvaceous hosts but can infect other species. In contrast, the DNA β molecules of the “non-Malvaceae type” tend not to be identified in malvaceous hosts [5].
Acknowledgements

M.S.S. is supported by a Ph.D. fellowship from Higher Education Commission (HEC), Government of Pakistan. R.W.B. is supported by the HEC under the ‘Foreign Faculty Hiring Program’.

References

Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal’s style. Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

**Please note**

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI]. If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.springerlink.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.
**Metadata of the article that will be visualized in OnlineFirst**

<table>
<thead>
<tr>
<th>Article Title</th>
<th>Characterization of begomovirus components from a weed suggests that begomoviruses may associate with multiple distinct DNA satellites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article Sub-Title</td>
<td></td>
</tr>
</tbody>
</table>
| Article CopyRight | Springer Science+Business Media, LLC  
(This will be the copyright line in the final PDF)                                                                 |
| Journal Name | Virus Genes                                                                                                                      |
| Corresponding Author | Family Name | Mansoor  
Particle |  |
| Given Name | S.  
Suffix |  |
| Division | Agricultural Biotechnology Division  
Organization | National Institute for Biotechnology and Genetic Engineering (NIBGE)  
Address | Jhang Road, P. O. Box 577, Faisalabad, Pakistan  
Email | shahidmansoor7@gmail.com |
| Author | Family Name | Mubin  
Particle |  |
| Given Name | M.  
Suffix |  |
| Division | Agricultural Biotechnology Division  
Organization | National Institute for Biotechnology and Genetic Engineering (NIBGE)  
Address | Jhang Road, P. O. Box 577, Faisalabad, Pakistan  
Email |  |
| Author | Family Name | Shahid  
Particle |  |
| Given Name | M. S.  
Suffix |  |
| Division | Agricultural Biotechnology Division  
Organization | National Institute for Biotechnology and Genetic Engineering (NIBGE)  
Address | Jhang Road, P. O. Box 577, Faisalabad, Pakistan  
Email |  |
| Author | Family Name | Tahir  
Particle |  |
| Given Name | M. N.  
Suffix |  |
| Division | Agricultural Biotechnology Division  
Organization | National Institute for Biotechnology and Genetic Engineering (NIBGE)  
Address | Jhang Road, P. O. Box 577, Faisalabad, Pakistan  
Email |  |
| Author | Family Name | Briddon  
Particle |  |
| Given Name | R. W.  |  |
A begomovirus disease complex associated with *Sonchus arvensis*, a common weed in Pakistan was studied using cloning, nucleic acid sequencing and phylogenetic analysis. The complex associated with this weed consists of a monopartite begomovirus and several distinct betasatellites and alphasatellites. The monopartite begomovirus associated with yellow vein disease of *Sonchus arvensis* showed 95–99% nucleotide sequence identity with *Alternanthera yellow vein virus* (AIYVV) reported from China, Vietnam and India. Two betasatellites were isolated from *S. arvensis*: one sharing between 91.4 and 95.3% nucleotide sequence identity with isolates of *Ageratum yellow leaf curl betasatellite* (AYLCB), and the other sharing between 78.2 and 99.9% identity with isolates of *Cotton leaf curl Multan betasatellite* (CLCuMB). Two alphasatellites were identified: one was homologous to *Potato leaf curl alphasatellite* (PotLCuA), while the other was closely related to *Hibiscus leaf curl alphasatellite* (HiLCuA). Thus, AIYVV in *S. arvensis* is associated with satellites shown previously to be associated with other begomoviruses in Pakistan. Our results suggest that monopartite begomoviruses may associate with distinct satellites that are prevalent in the region.
Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the ‘Author’s response’ area provided below.

<table>
<thead>
<tr>
<th>Query</th>
<th>Details required</th>
<th>Author’s response</th>
</tr>
</thead>
<tbody>
<tr>
<td>References [10, 11] are given in list but not cited in text. Please cite in text or delete from list.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Characterization of begomovirus components from a weed suggests that begomoviruses may associate with multiple distinct DNA satellites

M. Mubin · M. S. Shahid · M. N. Tahir · R. W. Briddon · S. Mansoor

Received: 5 January 2010 / Accepted: 2 March 2010
© Springer Science+Business Media, LLC 2010

Abstract A begomavirus disease complex associated with Sonchus arvensis, a common weed in Pakistan was studied using cloning, nucleic acid sequencing and phylogenetic analysis. The complex associated with this weed consists of a monopartite begomovirus and several distinct betasatellites and alphasatellites. The monopartite begomovirus associated with yellow vein disease of Sonchus arvensis showed 95–99% nucleotide sequence identity with Alternanthera yellow vein virus (AIYVV) reported from China, Vietnam and India. Two betasatellites were isolated from S. arvensis: one sharing between 91.4 and 95.3% nucleotide sequence identity with isolates of Ageratum yellow leaf curl betasatellite (AYLCB), and the other sharing between 78.2 and 99.9% identity with isolates of Cotton leaf curl Multan betasatellite (CLCuMB). Two alphasatellites were identified: one was homologous to Potato leaf curl alphasatellite (PotLCuA), while the other was closely related to Hibiscus leaf curl alphasatellite (HiLCuA). Thus, AlYVV in S. arvensis is associated with satellites shown previously to be associated with other begomoviruses in Pakistan. Our results suggest that monopartite begomoviruses may associate with distinct satellites that are prevalent in the region.

Keywords Sonchus arvensis · Geminiviruses · Begomoviruses · Alphasatellite · Betasatellite

Introduction

Geminiviruses are single-stranded (ss)DNA viruses with circular genomes that infect plants and are transmitted by insect vectors. Based upon their host range, nature of the insect vector and genome arrangement the geminiviruses have been classified into four genera: Begomovirus, Curtovirus, Mastrevirus and Topocuvirus. Geminiviruses transmitted by the whitefly, Bemisia tabaci, are classified in the genus Begomovirus [1]. The genomes of begomoviruses consist of either two genomic components of about equal size (~2,800 bp), known as DNA A and DNA B, or of a single component homologous to the DNA A component of the bipartite viruses [2]. The two components share little sequence similarity with the exception of an approximately 200 bp stretch of sequence with high nucleotide sequence identity (80–100%) that contains the origin of virion-strand DNA replication. The genomes of monopartite begomoviruses and DNA A components of bipartite begomoviruses encode all viral functions required for virus replication, control of gene expression and insect transmission, whereas the DNA B component encodes two proteins required for movement and symptom development in plants [3].

Recently, the majority of the begomoviruses originating from the Old World have been shown to be monopartite and to associate with a class of ssDNA satellites known as betasatellites (earlier known as DNA β) [4]. Betasatellites are approximately half the size of their helper begomoviruses (~1.4 kb) and are required by the helper virus to induce typical disease symptoms in their original hosts. The success of begomovirus–betasatellite disease complexes appears to be due to the promiscuous nature of betasatellites that allows them to be trans-replicated by several distinct begomoviruses [5–7].
betasatellite disease complexes are widespread throughout the Old World and outnumber bipartite begomoviruses whereas in the New World only bipartite begomoviruses are native. There have been recent reports which show that betasatellite can complement the function of DNA B, suggesting that the satellite may provide movement functions to its helper begomovirus [8]. In addition to betasatellites, many begomovirus—betasatellite disease complexes are also associated with a third single-stranded DNA component for which the collective term alphasatellite has been proposed (earlier known as DNA 1; R.W. Briddon, manuscript in preparation). However, alphasatellites are dispensable for virus infection and appear to play no significant role in the etiology of the diseases with which they are associated [9].

Geminiviruses are a group of rapidly emerging plant viruses. This can be attributed to various factors, including increased insect vector populations, the presence of alternative hosts and/or increasing spread due to international trade. Geminiviruses have the capacity to evolve rapidly in response to changes in their environment (such as alterations in cropping systems and/or population dynamics of insect vectors). Begomoviruses replicate in the nucleus and it appears that multiple begomovirus infections facilitate evolution of begomoviruses through pseudo-recombination and recombination [8]. Several distinct begomoviruses were found associated with the cotton leaf curl disease complex and suggest that distinct begomoviruses with the ability to interact with Cotton leaf curl Multan betasatellite (CLCuMB) were mobilized from distinct hosts such as papaya and tomato into cotton [5, 9]. Similarly, under experimental conditions, begomoviruses isolated from distinct geographical locations have the capacity to interact with betasatellites [6, 7, 12].

Weeds are reservoirs of begomoviruses that infect crop plants and act as “melting pots” that yield new viruses/virus strains by recombination and component exchange due to their frequently harboring multiple viruses. Nevertheless, weeds have in the past been neglected in the study of diversity of plant viruses. Weeds acting as reservoirs can play an important part in the emergence of plant viral epidemics affecting crops. Due to the importance of weed hosts in the epidemiology and evolution of begomoviruses, we have initiated a project to investigate the diversity of genomic components associated with weed hosts and have recently reported the begomoviruses and satellites associated with one such weed [13].

Sonchus arvensis (family Asteraceae) is a perennial weed species widely distributed in irrigated areas of Pakistan, especially around water channels, and frequently exhibits vein yellowing symptoms which are typical of begomoviruses. Here, we have characterized components of a begomovirus disease complex associated with this weed that consists of a monopartite begomovirus along with multiple beta- and alphasatellites.

Materials and methods

Collection of samples

During cotton-growing season of 2007, five samples of S. arvensis showing yellow vein symptoms were collected from cotton fields in the vicinity of the town of Samundri, district Faisalabad, Punjab province, Pakistan (Fig. 1).

Cloning and sequencing

Total DNA was extracted from symptomatic weed samples using a CTAB method [14]. Rolling circle amplification using φ DNA polymerase (Fermentas, Arlington, Canada) was performed to amplify all circular DNA molecules from DNA samples as described previously [15]. Universal primers were used to PCR amplify full-length begomovirus, betasatellite and alphasatellite [9, 16, 17] using concatameric φ polymerase product as the template. Amplification products of the expected sizes were cloned into the pTZ57R/T vector (Fermentas). Potentially full-length clones were sequenced commercially by primer walking (Macrogen, Korea).

Sequence analysis

Sequences were assembled and analyzed with the aid of the Lasergene package of sequence analysis software (DNA Star Inc., Madison, WI, USA). Multiple sequence alignments were performed using Clustal X [18]. Phylogenetic trees were constructed using Clustal X (neighbor-joining method), displayed, manipulated and printed using Treeview [19].

Fig. 1 Sonchus arvensis, a common weed showing characteristic vein yellowing symptoms of begomovirus infection
Results

*Alternanthera yellow vein virus* is associated with yellow vein disease of *Sonchus arvensis*

The association of a begomovirus with symptomatic *S. arvensis* was confirmed by Southern blot hybridization using *Cotton leaf curl Multan virus* as a general probe. All symptomatic samples hybridized with the probe yielding the bands characteristic of geminivirus replication (data not shown). PCR amplification with begomovirus specific primers yielded a single product of the expected size (~2.8 kb) which was cloned into the pTZ57R/T vector.

Seven clones were obtained which all showed identical restriction patterns and partial sequencing revealed them to have the highest levels of sequence identity to *Alternanthera yellow vein virus* (AlYVV) [20]. One clone was selected for complete sequencing. The complete sequence of the begomovirus was determined to be 2,748 nucleotides and is available in the databases under the accession number FN432361. Sequence comparisons showed the begomovirus to be closely related to isolates of AlYVV (99.7 to 99.3% nucleotide sequence identity to 9 AlYVV sequences available in the databases) with the highest to an isolate from China (isolate F30; accession number EU286797). This indicates that the virus isolated from *S. arvensis* is an isolate of AlYVV for which we propose the isolate descriptor [Pakistan:Samundri:Sonchus:2007] (AlYVV-[PK:Sam:Son:07]).

The virus cloned from *S. arvensis* has the typical genome organization of monopartite begomoviruses (or DNA-1 component of bipartite begomoviruses) with two ORFs in the virion-sense (encoding the V2 protein and coat protein [CP]) and four in the complementary sense (encoding the replication associated protein [Rep], the transcriptional activator protein [TrAP], the replication enhancer protein [Ren] and the C4 protein). Each of the ORFs show greater than 95% amino acid sequence identity to the homologous ORFs of AlYVV isolate F30 (Table 1) ruling out any possible occurrence of recombination.

A phylogenetic tree, based on an alignment of the complete nucleotide sequence of the begomovirus isolated from *S. arvensis* with selected begomovirus genome (or DNA-1 component) sequences is shown in Fig. 2A. This shows the sequence isolated from *S. arvensis* to segregate with isolates of AlYVV, confirming it identification as an isolate of this species.

Two distinct betasatellites are associated with the disease

The complete nucleotide sequences of the two betasatellite clones (MB05 and MB06) were determined to be 1,358 and 1,351 bp, respectively. These sequences are available in the databases under accession numbers FN432358 and FN432359, respectively. All previously reported betasatellite molecules contain three conserved regions; a sequence conserved between all betasatellites known as the satellite conserved region (SCR), a region of sequence rich in adenine (A-rich) and a single open reading frame in the complementary-sense called fC1 [12]. These features were identified in both sequences (Table 2). MB05 and MB06 share only 54.4% nucleotide sequence identity and this thus indicates that two different species of betasatellite are infecting the same plant; the presently applicable species demarcation threshold for betasatellites being 78% [21]. Clone MB05 clone shows between 91.4 and 95.3% sequence identity with isolates Ageratum yellow leaf curl betasatellite (AYLCB; nine sequences available in the databases), with the highest identity to AYLCB-[PK:BA-H:97](AJ316031). Clone MB06 showed between 78.2 and 99.9% sequence identity to isolates of CLCuMB (85 sequences available in the databases) with highest identity to four isolates recently characterized from southern Pakistan (accession numbers AM712311, AM712314, AM712315, AM712320). MB06 lacks the approx. 200 nt recombinant fragment in the SCR derived from Tomato leaf curl betasatellite, marking it as an isolate of CLCuMB typical of CLCuD prevalent in cotton during the 1990 s rather than the recombinant CLCuMB now associated with the disease [22]. A phylogenetic tree, based on an alignment of the complete nucleotide sequence of the betasatellites isolated from *S. arvensis* with selected betasatellite sequences available in the databases is shown in Fig 2B. This shows MB05 and MB06 to segregate with isolates of AYLB and CLCuMB, respectively, confirming their placement within these two species. We propose the isolate descriptors AYLB-[Pakistan:Samundri:Sonchus:2007] (PK:SamMB05:Sonchus:07) and CLCuMB-[Pakistan:Samundri:MB06:Sonchus:2007] (PK:SamMB06:Sonchus:07) for these sequences.

Two distinct alphasatellites are associated with the disease

Six alphasatellite clones were completely sequenced and the sequences are available in the databases under accession numbers given in Table 3. The sequences showed the typical arrangement of alphasatellites, containing a single large open reading frame encoding a replication associated protein (Rep) in the virion-sense. The sequence of UK1 showed 99.3 to 99.9% nucleotide sequence identity with isolates Potato leaf curl alphasatellite (PotLCuA; two sequences available in the databases), an alphasatellite we earlier identified in potato and the weed *Digeria arvensis* associated with Chili leaf curl virus
Table 1 Genes encoded by the begomovirus isolated from S. arvensis and amino acid sequence identities to an isolate of AlYVV originating from China

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates</th>
<th>Coding capacity (no. of amino acids/predicted mol. weight [kDa])</th>
<th>Percentage amino acid sequence identity to AlYVV (EU286797)(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>314-1084</td>
<td>257/29.8</td>
<td>97.3</td>
</tr>
<tr>
<td>V2</td>
<td>154-495</td>
<td>114/12.8</td>
<td>95.6</td>
</tr>
<tr>
<td>Rep</td>
<td>2621-1539</td>
<td>361/41.1</td>
<td>99.7</td>
</tr>
<tr>
<td>TrAP</td>
<td>1633-1232</td>
<td>134/15.5</td>
<td>98.5</td>
</tr>
<tr>
<td>REn</td>
<td>1488-1087</td>
<td>134/15.9</td>
<td>99.3</td>
</tr>
<tr>
<td>C4</td>
<td>2470-2177</td>
<td>98/11.1</td>
<td>100</td>
</tr>
</tbody>
</table>

\(^a\) An isolate of AlYVV from China to which the begomovirus shows the highest levels of sequence identity.

Fig. 2 Phylogenetic dendrograms based upon the complete nucleotide sequences of selected begomoviruses (a), betasatellites (b) and alphasatellites (c). Vertical branches are arbitrary; horizontal branches are proportional to calculated mutation distance. Values at nodes indicate percentage bootstrap values (1,000 replicates). Begomovirus acronyms used are Alternanthera yellow vein virus (AlYVV), Bean golden yellow mosaic virus (BGYMV), Cabbage leaf curl Jamaica virus (CabLCuJV), Chili leaf curl virus (ChLCV), Cotton leaf curl Kokkhan virus (CLCuKv), Cotton leaf curl Multan virus (CLCuMB), Cowpea golden mosaic virus (CPGMV), Cucurbit leaf crumple virus (CuLCrV), Indian cassava mosaic virus (ICMV), Malvstrum leaf curl virus (MalLCV), Mungbean yellow mosaic India virus (MYMV), Okra yellow crinkle virus (OYCv), Papaya leaf curl China virus (PaLCuCNV), Pepper leaf curl Bangladesh virus (PeplCBVD), Pepper leaf curl Lahore virus (PeplC LV), Pepper yellow leaf curl Indonesia virus (PeplCIV), Radish leaf curl virus (RaLCV). The isolate descriptors used are as given in [28]. The betasatellite acronyms used are as given in [21]. Alphasatellite acronyms used are Tomato leaf curl China alphassatellite (TYLCNA), Tomato leaf curl Bangladesh betasatellite (ToLCBD), Cotton leaf curl Multan betasatellite (CLCuMB), Malvstrum yellow vein Yunnan betasatellite (MaYVYnB) and Bhendi yellow vein betasatellite (BYVB). The betasatellite isolate descriptors used are as given in [21]. Alphasatellite acronyms used are Tomato yellow leaf curl China alphassatellite (TYLCNA), Cotton leaf curl Multan alphassatellite (CLCuMA), Ageratum yellow vein alphassatellite (AYVA), Tomato leaf curl alphassatellite (ToLCA), Hibiscus leaf curl alphassatellite (HiLCuA), Okra leaf curl alphassatellite (OLCuA), Malvstrum yellow mosaic alphassatellite (MaYMA), Sida yellow vein Vietnam alphassatellite (SiYVV), Tobacco yellow leaf curl Yunnan alphassatellite (TYLCYnA) and Tobacco curly leaf alphassatellite (TbCSA). The trees were arbitrarily rooted on the sequence of Tomato leaf curl New Delhi virus DNA B (ToLNDV-DNA B [U15017] (for the begomovirus phylogenetic tree), Tomato yellow leaf curl Yunnan alphassatellite (TYLCYnA-[CN:Yn244:04]) (for the betasatellite phylogenetic tree) and Cotton leaf curl Multan betasatellite (CLCuMB-[PK-Fai:Tom:05]AM490309) (for the alphassatellite phylogenetic tree) which are unrelated sequences of a similar size. The database accession numbers are indicated in each case. The sequences originating from S. arvensis are indicated by a black box in each case.

252 and Cotton leaf curl Rajasthan virus, respectively [13, 23]. Clones KD5, UKV, UK7, UK8 and UK12 showed between 253 85.5 and 95.6% nucleotide sequence identity with isolates 254 of Hibiscus leaf curl alphassatellite (HiLCuA; 7 full-length 255 sequences available in the databases). A phylogenetic tree 256 based upon the full length sequences of the alphassatellites 257 characterized here with selected alphassatellite sequences 258 available in the databases is shown in Fig. 2c. This shows 259
The virus has also been identified in Vietnam and Pakistan. The trade by land route may have restricted by the Himalayan mountain range, which serves as a natural geographical barrier. In contrast, the viruses found in India, Pakistan and Bangladesh are often related due to their geographical proximity with no natural barriers. This is the first report of a begomovirus found in China, Vietnam and Pakistan. The trade by land route may have disseminated the virus in China, Pakistan and India.

### Table 2 Features of the betasatellites isolated from S. arvensis

<table>
<thead>
<tr>
<th>Clone</th>
<th>Coordinates/no. of amino acids</th>
<th>Predicted coding capacity [kDa]</th>
<th>SCR (coordinates)</th>
<th>A-rich (coordinates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MB01</td>
<td>596-183/138</td>
<td>16.0</td>
<td>1253-18</td>
<td>694-977</td>
</tr>
<tr>
<td>MB02</td>
<td>551-198/118</td>
<td>13.7</td>
<td>1246-18</td>
<td>766-972</td>
</tr>
</tbody>
</table>

### Table 3 Features of the alphasatellites isolated from S. arvensis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KD5</td>
<td>[PK:KD5:Son:08]</td>
<td>AM884370</td>
<td>1375</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.5</td>
<td>88.3–95.6</td>
</tr>
<tr>
<td>UKV</td>
<td>[PK:UKV:Son:08]</td>
<td>AM930244</td>
<td>1374</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.5</td>
<td>85.5–95.3</td>
</tr>
<tr>
<td>UK12</td>
<td>[PK:UK12:Son:08]</td>
<td>AM930245</td>
<td>1375</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.5</td>
<td>88.2–95.6</td>
</tr>
<tr>
<td>UK7</td>
<td>[PK:UK7:Son:08]</td>
<td>AM930246</td>
<td>1375</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.5</td>
<td>86.0–95.6</td>
</tr>
<tr>
<td>UK8</td>
<td>[PK:UK8:Son:08]</td>
<td>AM930248</td>
<td>1374</td>
<td>78-1025/948</td>
<td>315</td>
<td>36.5</td>
<td>86.0–95.3</td>
</tr>
<tr>
<td>UK1</td>
<td>[PK:UK1:Son:08]</td>
<td>FN432360</td>
<td>1369</td>
<td>74-1015/942</td>
<td>314</td>
<td>36.3</td>
<td>69.9–65.1</td>
</tr>
</tbody>
</table>

Note: Number of full-length sequences available in the databases

The identification of the initial viral inoculum source is important for the successful control of plant viral diseases. S. arvensis is a perennial weed and is found commonly around water channels and in crop fields. We identified a monopartite begomovirus associated with multiple betasatellite and alphasatellite molecules infecting this weed. AlYVV was first identified in China and was identified as a natural geographical barrier. In contrast, Ha et al. [26] showed the AlYVV isolate infecting Zinnia elegans, originating from Vietnam, to be associated with a betasatellite that was later named Alternanthera yellow vein betasatellite [21]. It would thus appear that AlYVV has the capacity to cause disease, in the field, in either the presence or absence of a betasatellite.

Recombination among begomoviruses and component capture contribute to emergence of new complexes, and our earlier findings have shown that a recombinant betasatellite is associated with resistance breakdown in cotton [22]. We have also identified multiple and recombinant betasatellite molecules in another weed host [13]. It appears that weeds are likely vessels for mobilization and recombination due to multiple infections of begomoviruses and their associated satellites. The presence in S. arvensis of two betasatellites, including the betasatellite shown to be associated with CLCuD, and two distinct alphasatellites highlights this possibility. The presence of alphasatellites in weeds is significant because their presence may possibly ameliorate symptoms and possibly reduce virus titre which is essential for the successful propagation of virus infection [27]. The demonstration of the wide diversity of viruses and satellites in weed hosts illustrates the potential importance of weeds in the emergence of virus diseases of crops in the...
Indian subcontinent and other parts of the Old World. We suggest that AlYVV is probably an ancient virus that evolved in weeds that has the capacity to interact with diverse DNA satellites. Our data further show that monopartite begomoviruses can associate with diverse DNA satellites that may facilitate their mobilization into diverse hosts. These results should serve as the basis for further analysis of symptomatic and non-symptomatic weeds that are likely reservoirs that can harbor multiple begomoviruses may lead to the generation of new species/strains by recombination and component exchange.

Acknowledgements MM, MSS, and MNT are supported by Ph.D. fellowships from Higher Education Commission (HEC), Government of Pakistan. RWB is supported by the HEC under the “Foreign Faculty Hiring Program”. This research has been supported by projects from the Ministry of Science and Technology, Government of Pakistan, and a project between NIBGE and the Donald Danforth Plant Science Center (St. Louis, USA) under the Pakistan-USA linkage program.

References